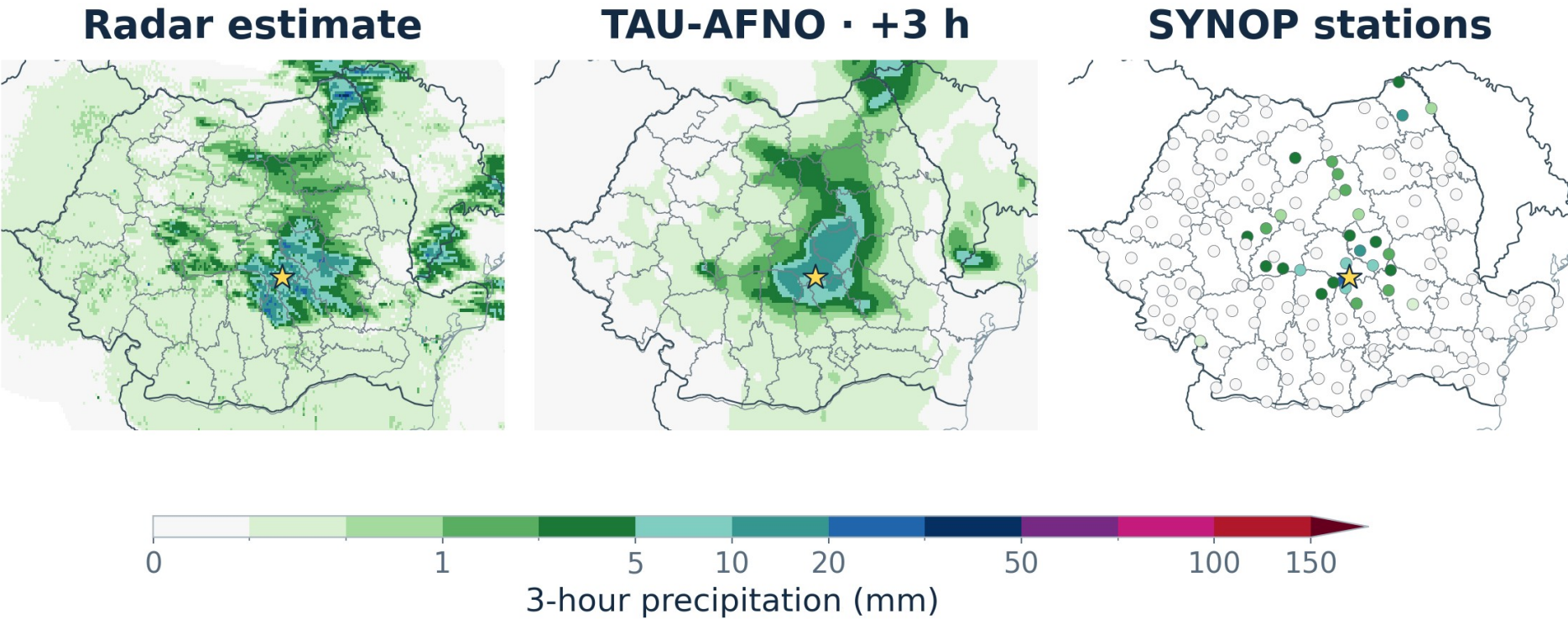


WINDMIL-AI: research to pre-production

Ioan Stefan Gabrian

National Meteorological Administration (NMA), Romania



26 August 2026 · 17–20 UTC | ★ Predeal: 36.8 mm observed in 3 h

28th COSMO General Meeting · Rome, Italy · 14–18 September 2026

Trained AI models at the core of WinDMiL

INSTITUTIONAL COORDINATOR • UBB

Gabriela Czibula

Babeș-Bolyai University

INSTITUTIONAL COORDINATOR • NMA

Ioan-Ștefan Gabrian

National Meteorological Administration, Romania

Intelligent systems & methods • Software development & engineering

Operational & research meteorology

Main product • A working ML decision-support prototype

NMA, direct beneficiary • Faster decision times and higher precision

O1 • Develop, train and scientifically validate

ML models learning historical weather-evolution patterns through radar-satellite data fusion.

Forecast weather maps and parameters from remote-sensing data.

O2 • Integrate and evaluate with users

Trained models + descriptive / predictive analysis + meteorologist-readable visualisation.

Quantitative precipitation forecasting; severe weather (rainfall, hail).

Funding: PN-IV-P7-7_1-PED-2024-0121 • Contract 13PED/2025

cs.ubbcluj.ro/windmil

The WinDMiL project team

Babeș-Bolyai University

UBB

Gabriela Czibula

Professor, PhD • UBB coordinator

Istvan Czibula

Professor, PhD

Arthur Molnar

Associate Professor, PhD

Mihai Andrei

Lecturer, PhD

Mariana Maier

Lecturer, PhD

National Meteorological Administration

NMA • Romania

Ioan-Ștefan Gabrian

Engineer, PhD student • NMA coordinator

Eugen Mihuleț

Forecaster

Raluca Gabrian

Engineer, PhD student

Meda Daniela Andrei

Forecaster, PhD

AI & software engineering • Operational & research meteorology

cs.ubbcluj.ro/windmil

Research now supports an hourly forecasting chain

Targets & observations

Corrected TP/WW
BUFR, radar, FCI

Regional AI

U-Net → graph models
→ TAU-AFNO

Forecast guidance

Ensemble, calibration
TP + present weather

Verification

Stations + radar
Events, leads, thresholds

Pre-production

Hourly delivery
Recovery + provenance

NWP / global AI

ICON observation updates
ICON → WeatherNext 2

From image-based learning to a hybrid weather nowcaster

1 • U-Net

Image-to-image encoder-decoder

32 → 48 → 64 base channels
BC64 pre-production model

Multiscale image patterns
History represented as channels

Overfitting informed selection;
epoch 34 retained for TP.

2 • Graph models

GraphCast-style • Anemoi-aligned

H256 • 8 processor layers
H384 • 12 processor layers

Graph-based spatial relationships
Neighbour-to-neighbour messages

Complementary TP / WW results
Both retained with U-Net
in a lead-dependent ensemble.

3 • TAU-AFNO

Temporal Attention Unit +
Adaptive Fourier Neural Operator

H1024 • 12 TAU blocks
16 AFNO blocks
Per block: 16 × 64 channels

Explicit temporal attention
Domain-wide Fourier mixing

Hybrid nowcaster proposed
Epochs 4 / 9 / 16 retained
TP • WW • P10 / P50 / P90



Shared forecast design: 7 history frames • 138 × 288 grid • 13 direct leads from 0 to 6 hours

Research outcome: temporal attention + global Fourier mixing + local detail in one hybrid nowcaster.

Time-correct observations are part of the model

BUFR / SYNOP

- All station subsets
- Field-wise corrections
- Exact accumulation windows

Radar + FCI

- Seven aligned history times
- Radar geometry diagnostics
- Explicit missing-data masks

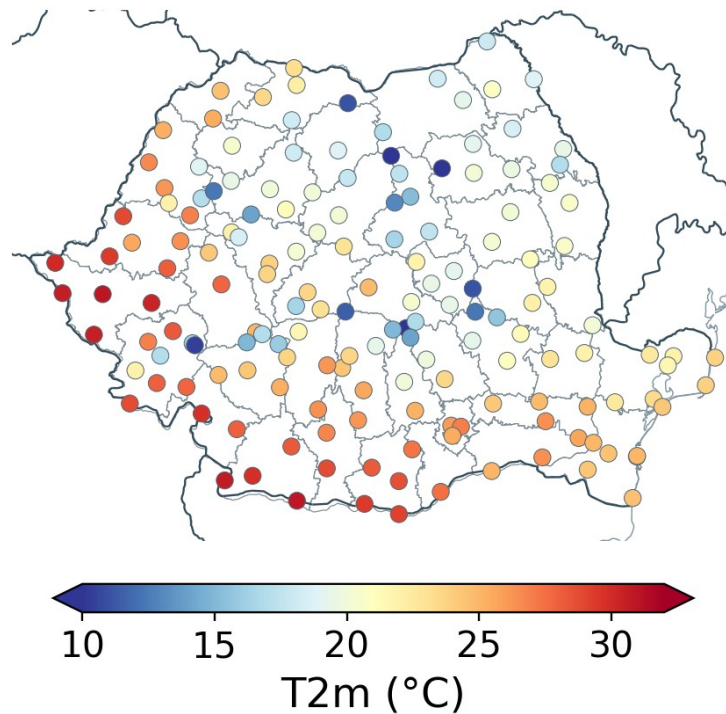
Retained result

- 156 stations in all 7 frames
- 30 July structural audit
- Radar QC kept in shadow

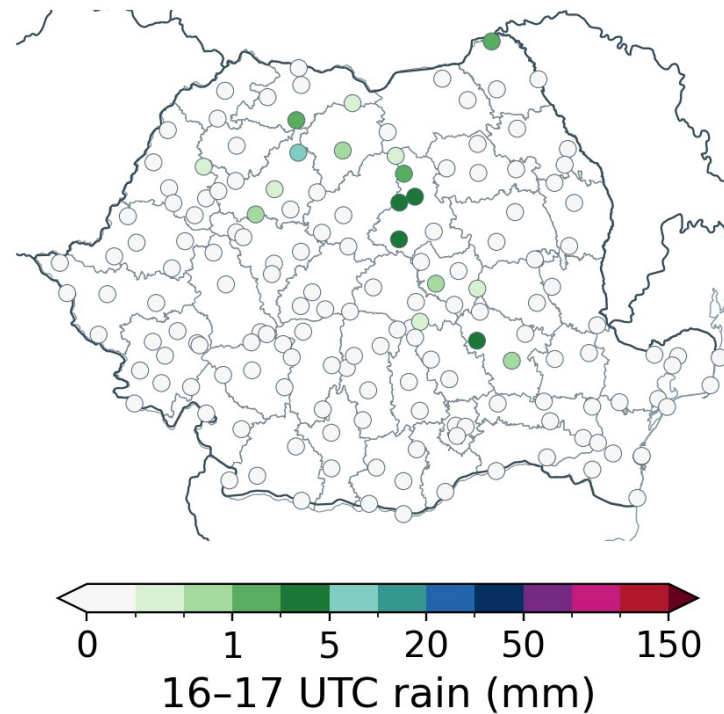
A forecast history must represent what was available at its own initialization time.

BUFR anchors the model to observed surface weather

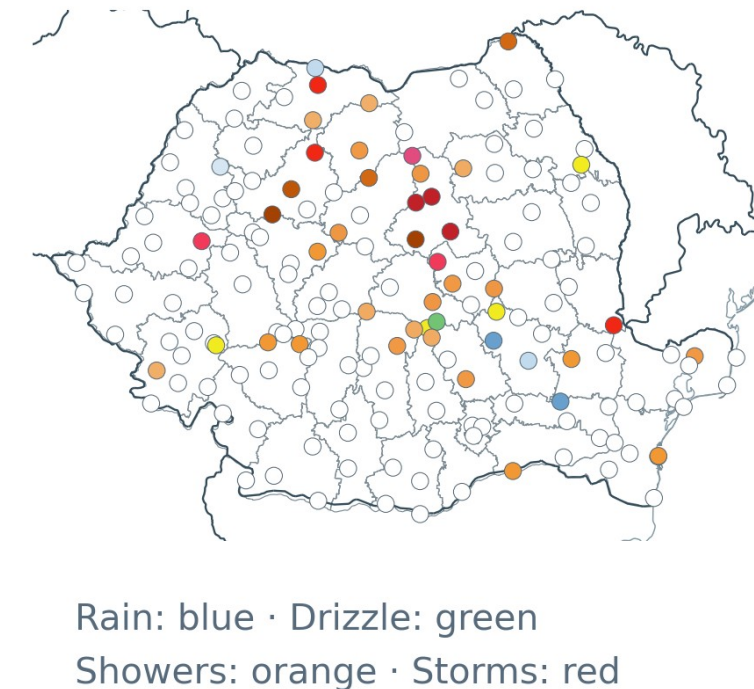
Surface temperature



Hourly precipitation



Present weather



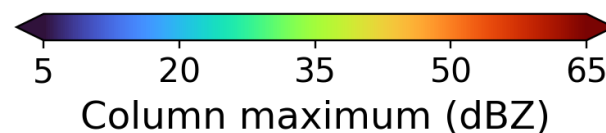
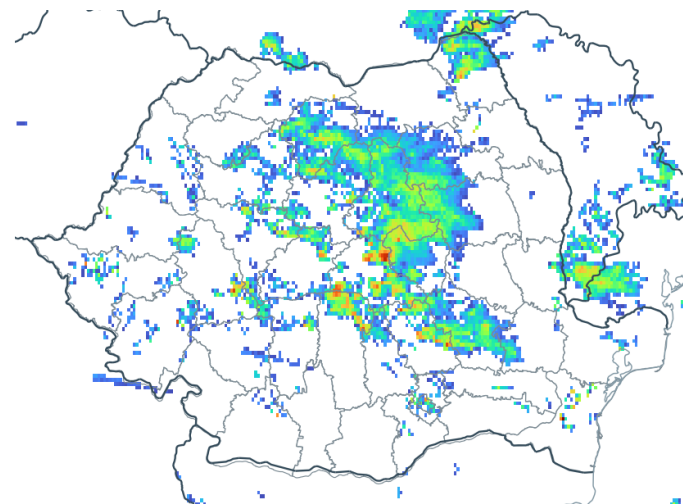
Inputs: rain, WW, temperature, humidity, wind, visibility and pressure

Training: real SYNOP constrains ICON-downscaled ERA5 TP/WW targets

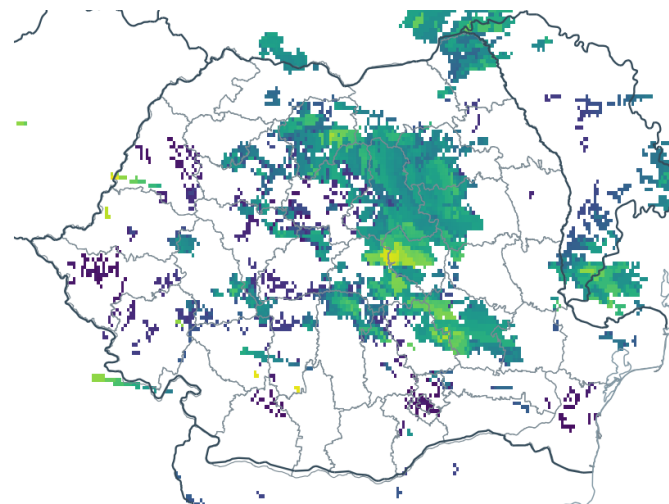
Station observations keep the learning objective tied to weather measured at the surface.

Radar supplies the vertical structure of precipitation

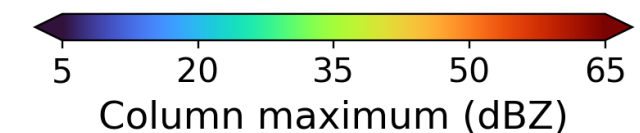
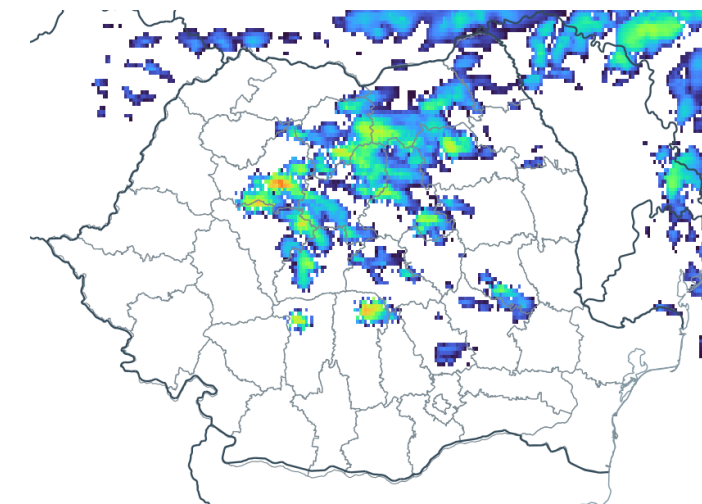
Observed radar
26 Aug 2026 • 17 UTC



Mapped echo height
26 Aug 2026 • 17 UTC



Simulated radar
26 Aug 2026 • 17 UTC



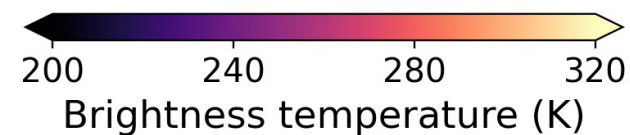
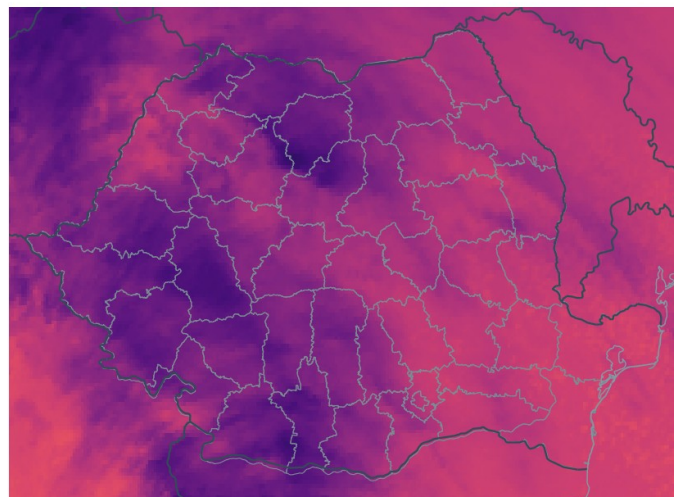
Pre-production: seven radar sites → 50 height-mapped reflectivity channels

Training: ICON-downscaled ERA5 → 50 simulated reflectivity levels

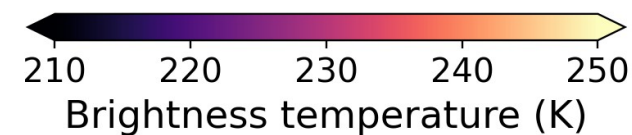
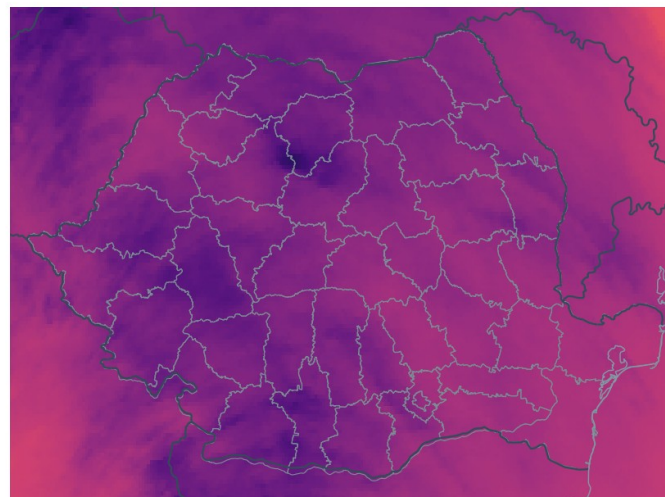
The network receives echo depth and spatial structure, not only a two-dimensional rain map.

Brightness temperatures add cloud and moisture structure

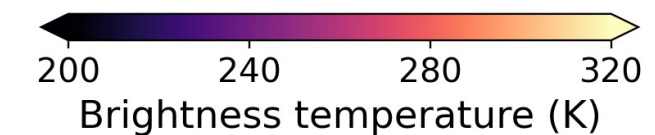
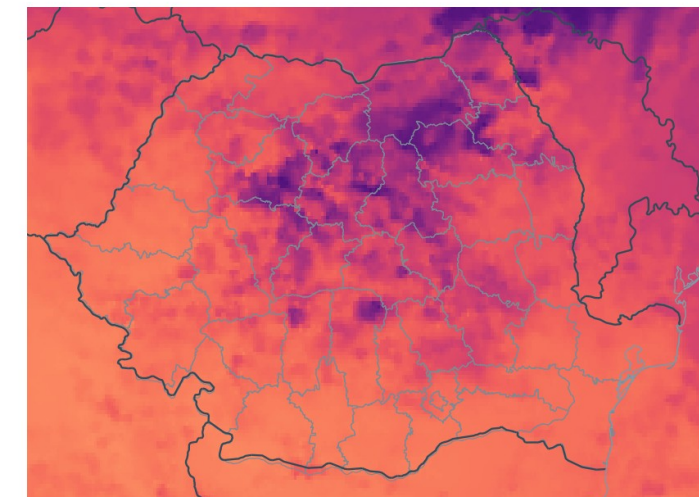
Observed FCI · IR12.3
26 Aug 2026 · 17 UTC



Observed FCI · WV6.3
26 Aug 2026 · 17 UTC



Synthetic · IR12.1*
26 Aug 2026 · 17 UTC

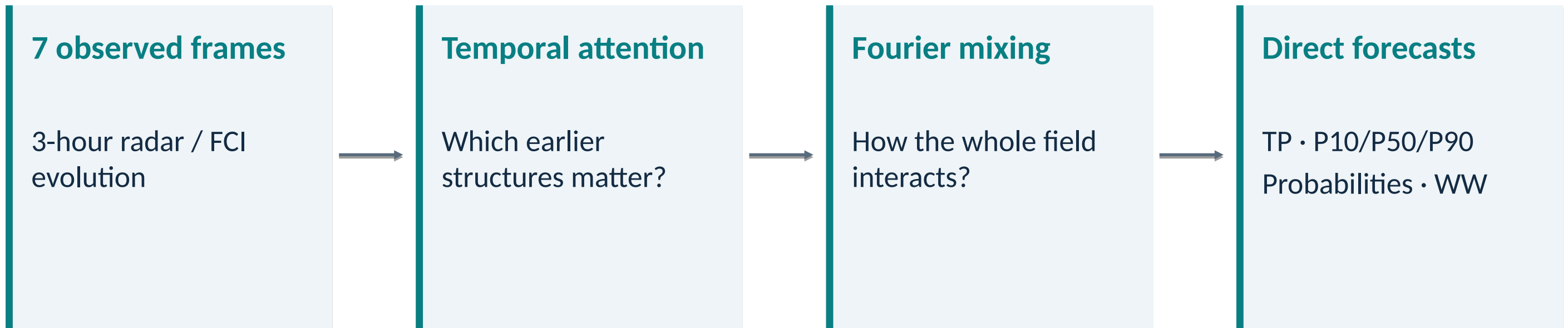


Pre-production: six FCI infrared / water-vapour channels → common model grid

Training: ICON-downscaled ERA5 → six synthetic BT channels

Cloud and moisture patterns complement radar. *IR12.1 is the stored training-channel name.

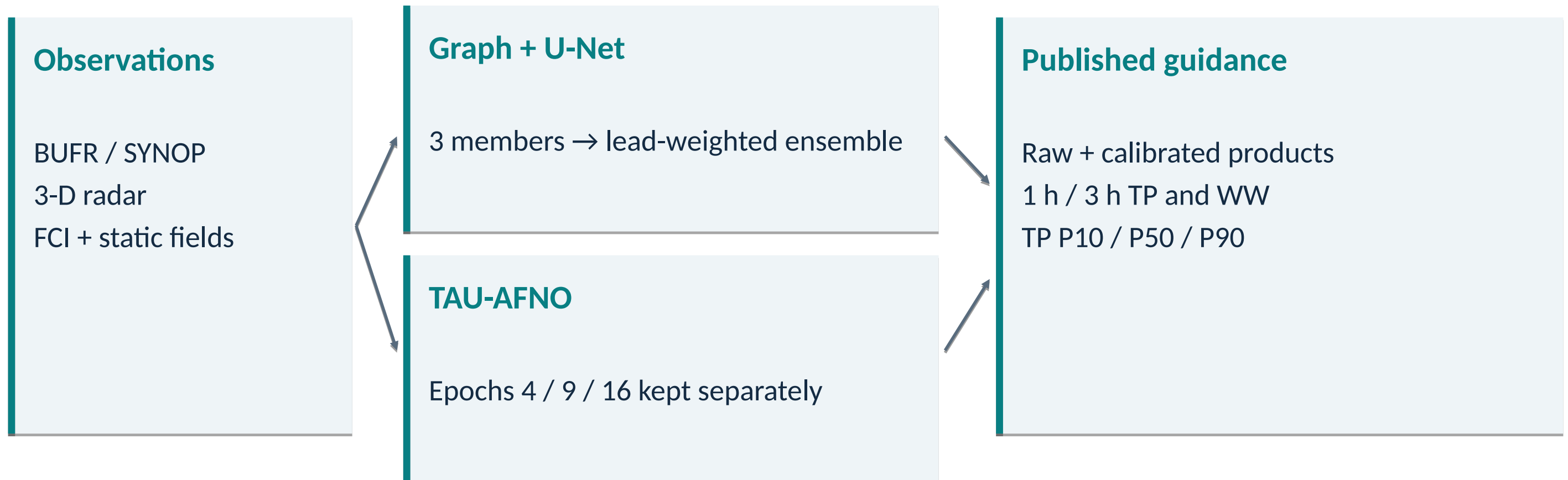
TAU-AFNO learns temporal and domain-wide structure



Station residual + bounded advection + local detail

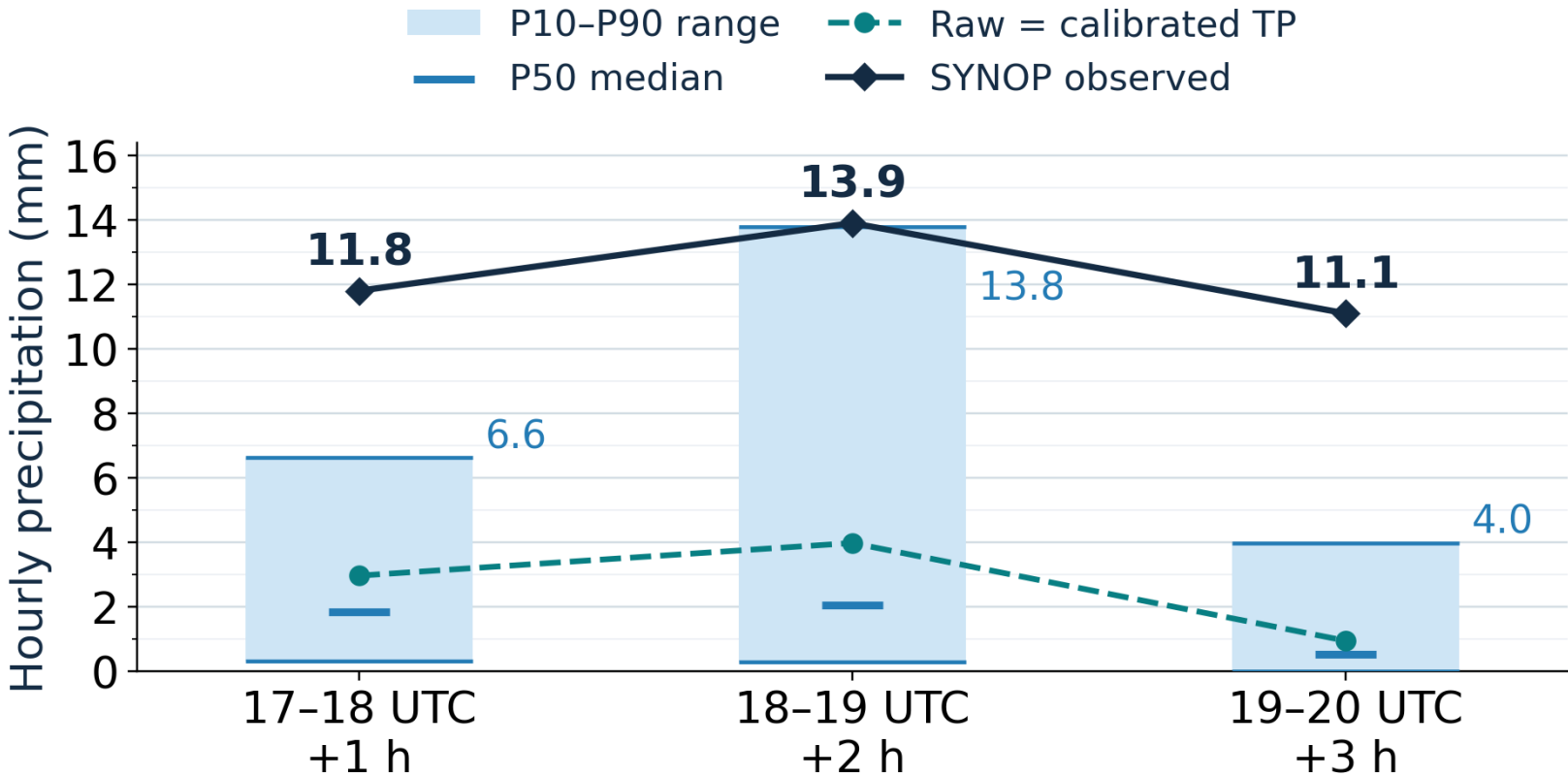
204.4 million parameters · 138×288 grid · 13 direct leads from 0 to 6 hours

Six neural members run in one pre-production chain



Raw forecasts, calibrated candidates and verification records retain separate identities.

Peak-hour P90 closely matches the observed rainfall



Raw hourly TP quantiles

P10 • lower

10th percentile

P50 • median

50th percentile

P90 • wetter

90th percentile

18-19 UTC • peak hour

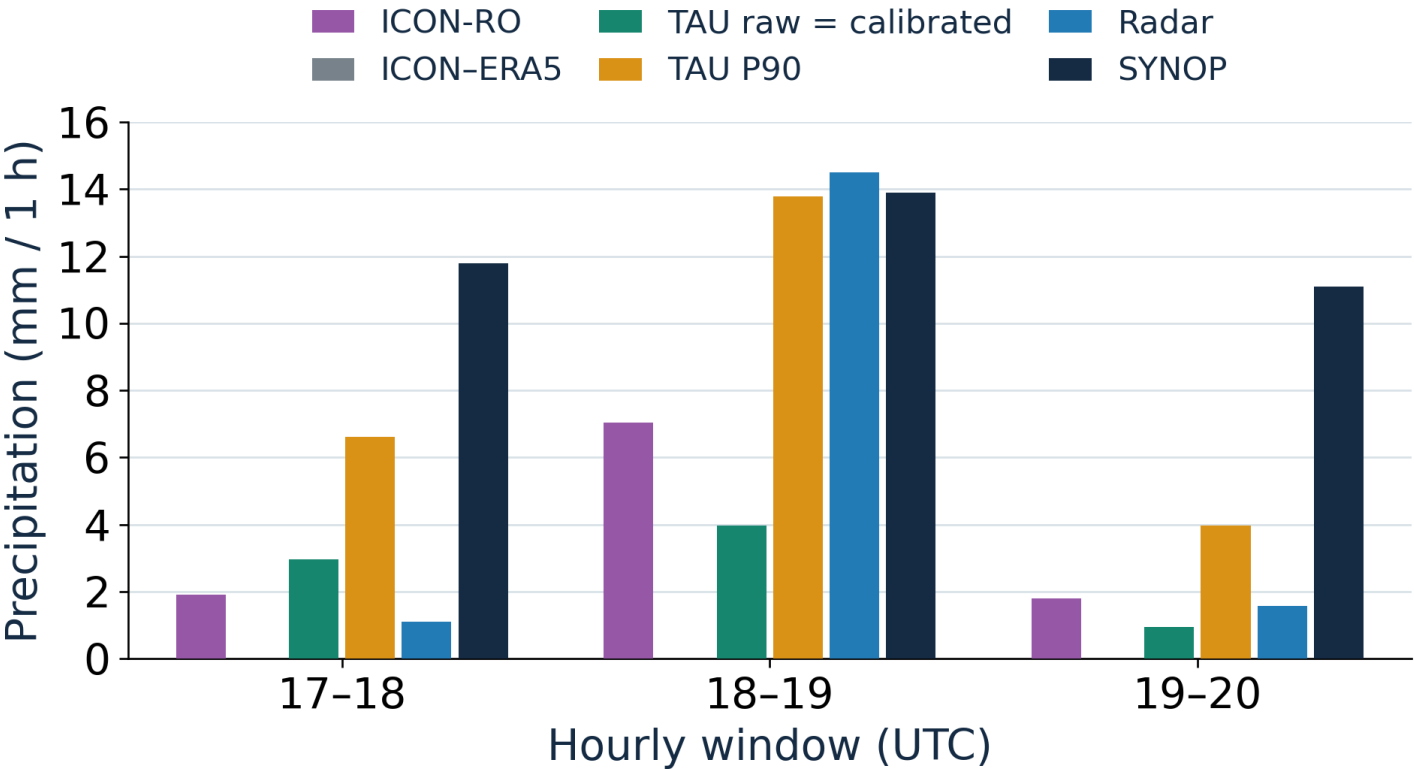
P90 13.8 • SYNOP 13.9 mm

3 h: raw = calibrated 7.9 mm • sum of hourly P90 24.4 mm* • SYNOP 36.8 mm

*Sum of hourly P90 values—not the 90th percentile of the 3 h total.

WW: 61 slight rain → 81 moderate/heavy showers; SYNOP 183 heavy showers.

Hourly TAU-AFNO guidance complements twice-daily ICON-RO forecasts



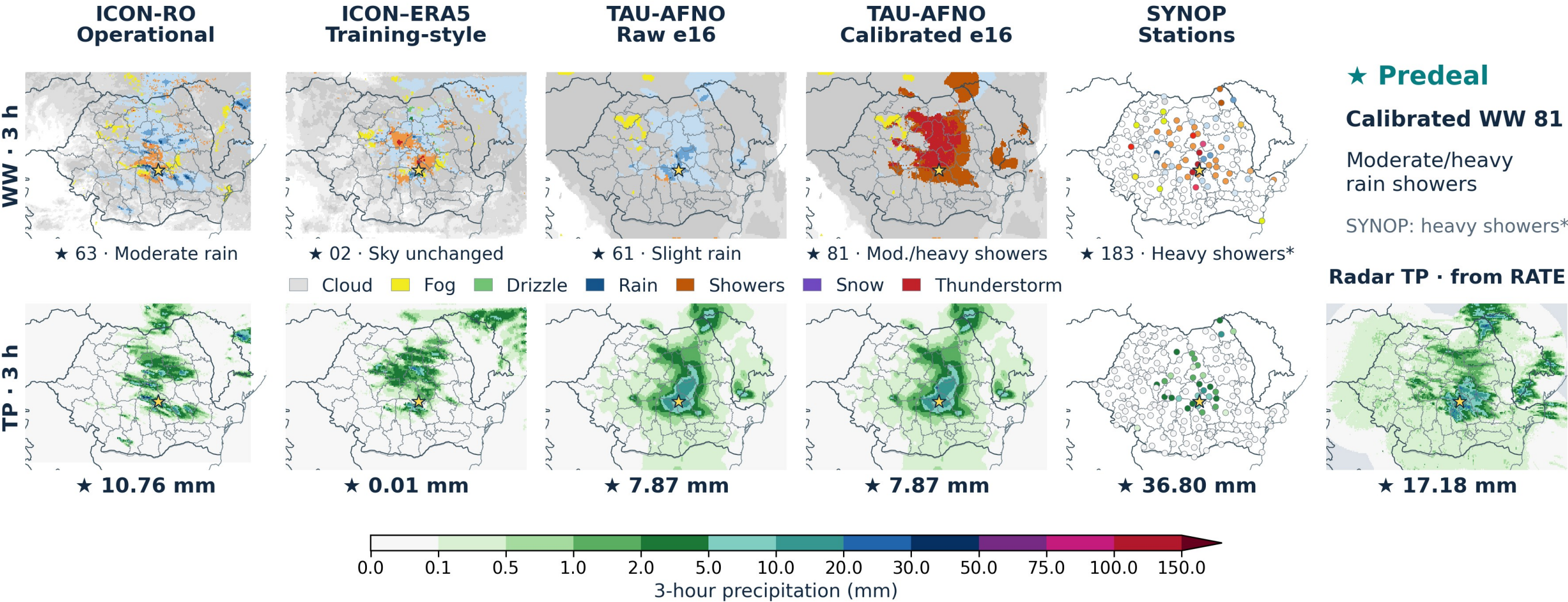
	WW at hour end		
	18 UTC	19 UTC	20 UTC
ICON-RO	61	63	61
ICON-ERA5	02	02	03
TAU raw	61	61	61
TAU calibrated	81	81	81
SYNOP	181	181	183

02 Sky unchanged • 03 Clouds increasing • 61/63 Slight/moderate continuous rain
81 Moderate/heavy rain showers • 181/183 Slight/heavy showers or intermittent rain (automatic)

Strengthening heavy-rain decision support: P90 reveals potential rainfall intensity, while calibrated weather codes identify expected phenomena—with forecasts generated in tens of seconds on CPU from prepared inputs.

Case cycles: ICON 00 UTC • TAU 17 UTC • P90: raw hourly quantile • Radar: not gauge-adjusted

National maps connect rainfall guidance with weather type



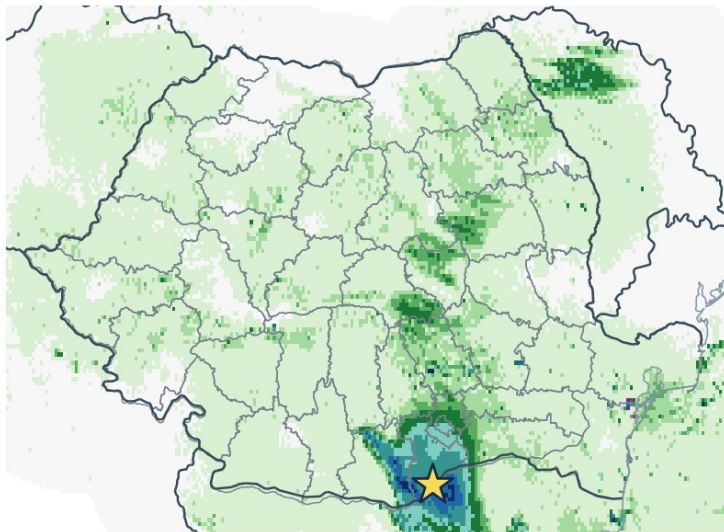
TP: 3-hour total · WW: most significant hourly endpoint code · * Showers or intermittent rain

Use rainfall maps for location and amount; calibrated WW adds the expected weather type.

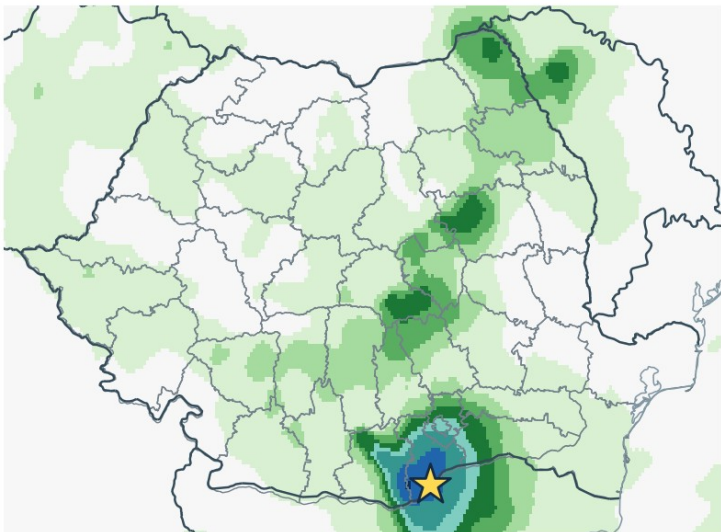
ICON cycles 00 UTC · TAU 17 UTC · Radar: not gauge-adjusted; grey cells indicate insufficient coverage

Giurgiu: 24.8 mm forecast, 19.1 mm observed

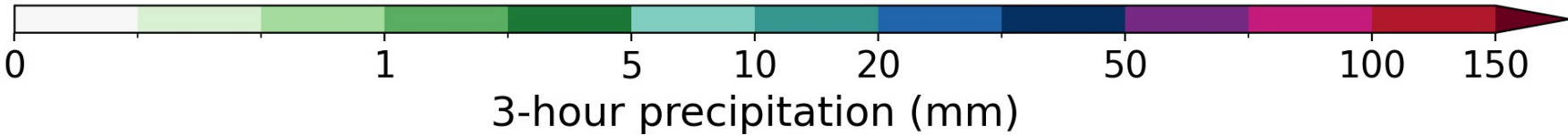
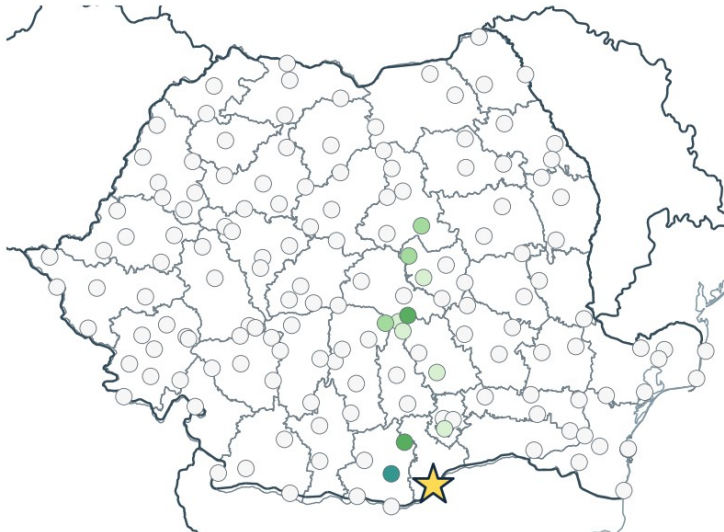
Radar estimate



TAU-AFNO · +3 h



SYNOP stations



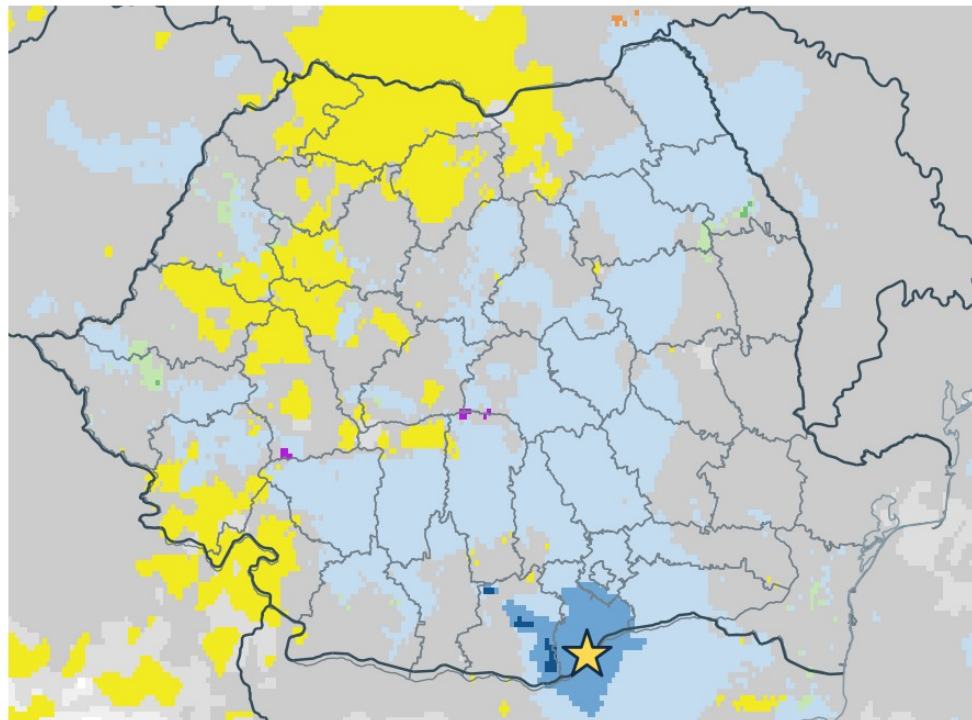
★ GIURGIU (15491)

Radar 18.5 mm · TAU 24.8 mm · SYNOP 19.1 mm

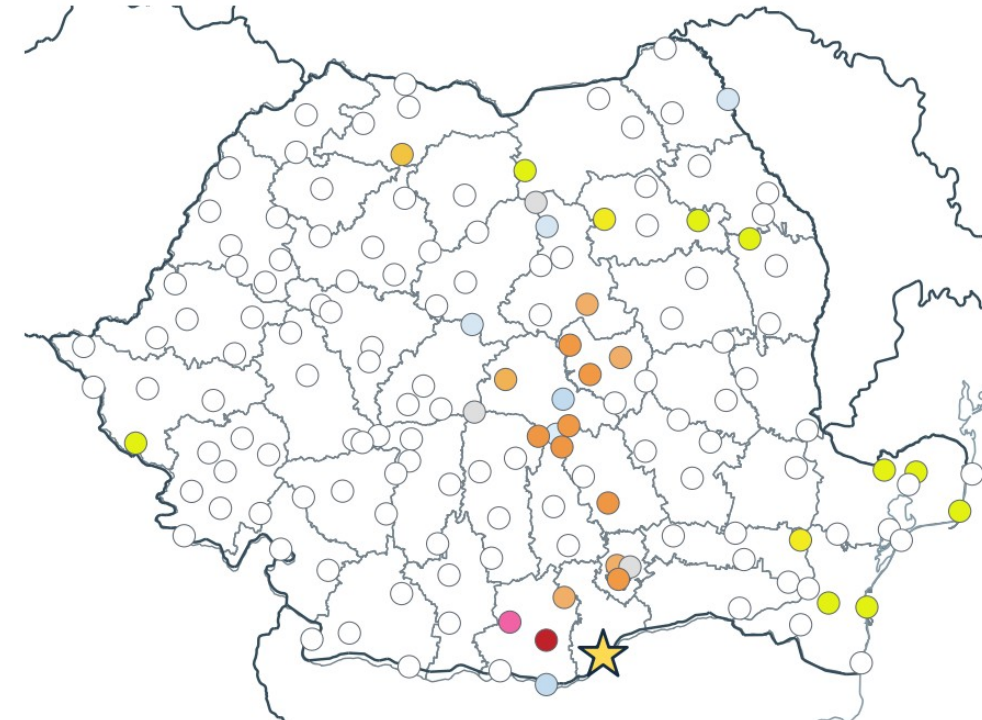
At the highest-rainfall station, TAU overestimates the three-hour total by 5.7 mm.

At Giurgiu, the forecast weather type differs from SYNOP

TAU-AFNO · +1 to +3 h



SYNOP · same 3 hours



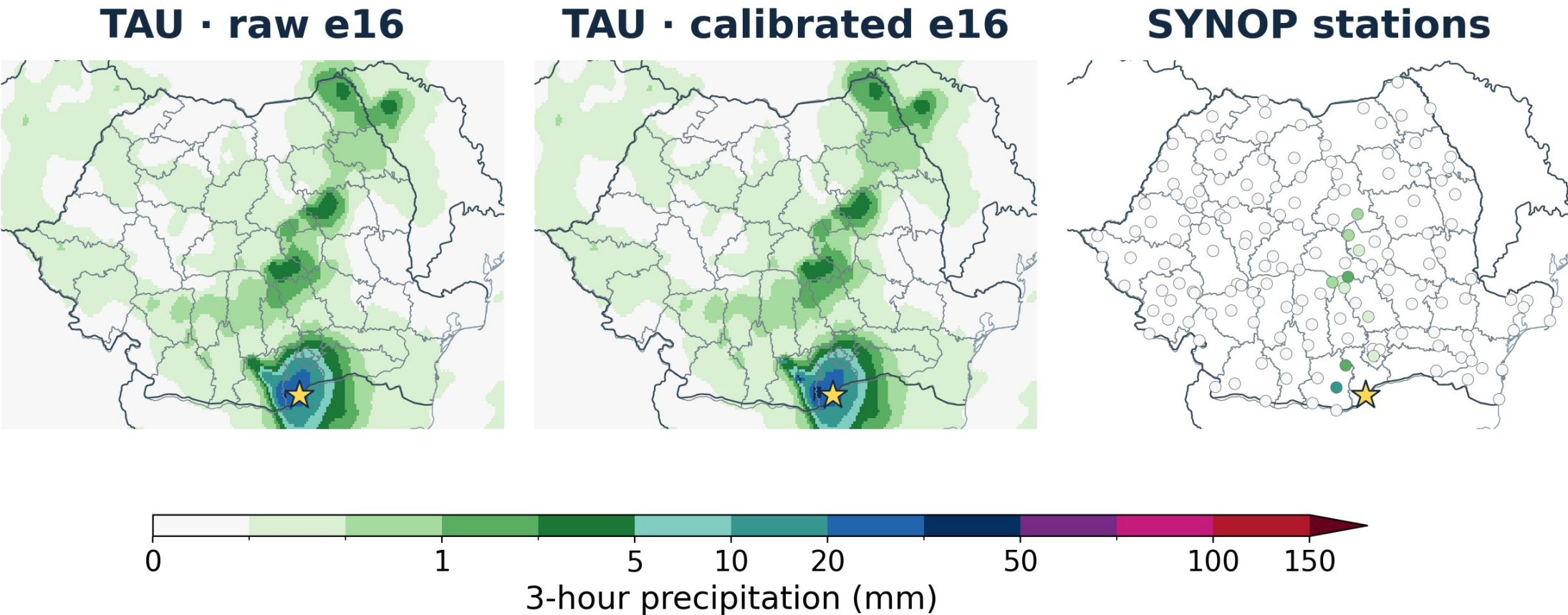
★ GIURGIU

TAU · WW 63: moderate continuous rain

SYNOP · WW 80: slight rain showers

Rain: blue · Drizzle: green · Showers: orange · Fog: yellow · Thunderstorms: red

Calibration increases the Giurgiu rainfall overestimate

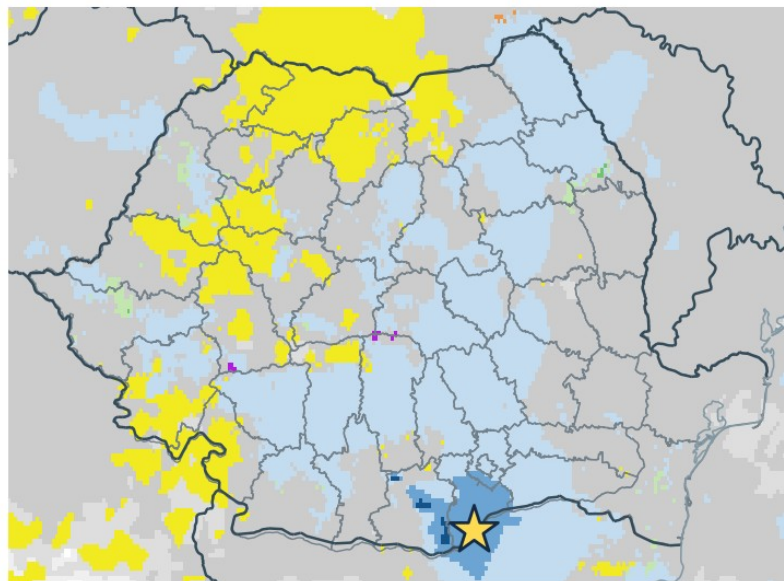


★ GIURGIU Raw 24.8 mm · Calibrated 28.1 mm · SYNOPSIS 19.1 mm

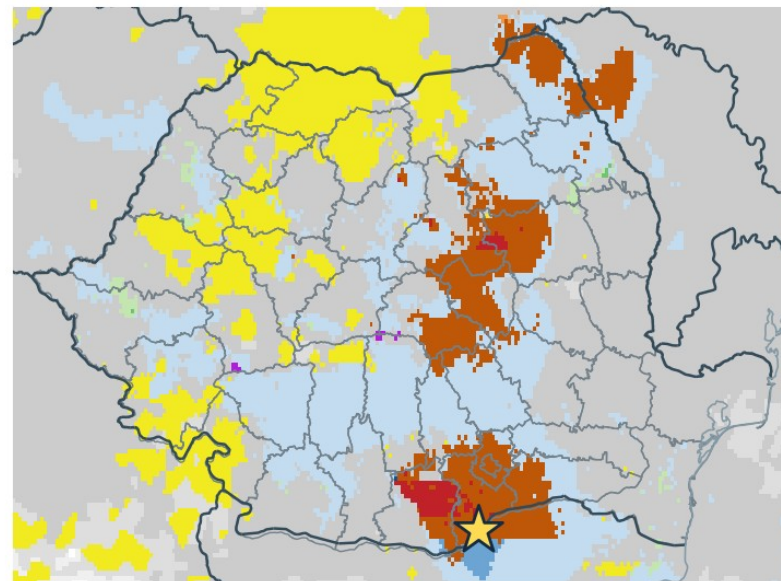
The three-hour station error grows from +5.7 to +9.0 mm; this layer cannot lower rainfall.

Calibration identifies showers, but overstates their intensity

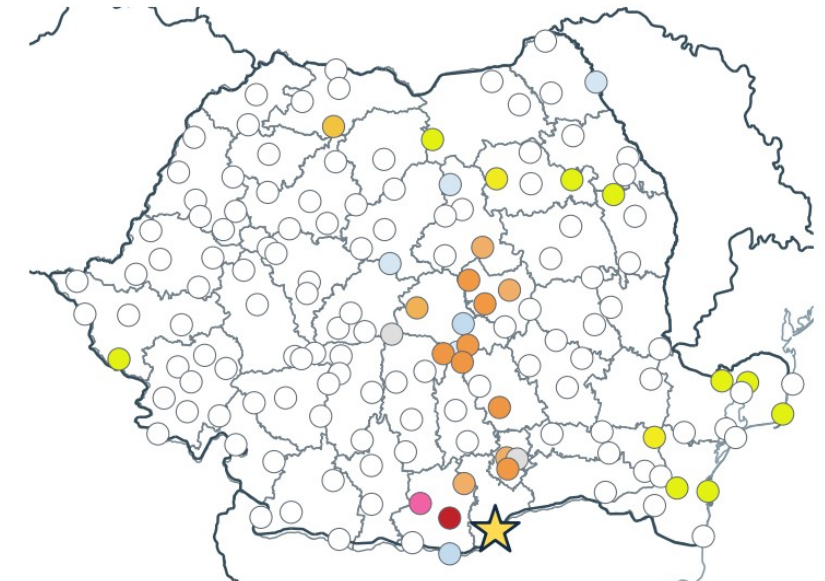
TAU · raw e16



TAU · calibrated e16



SYNOP stations



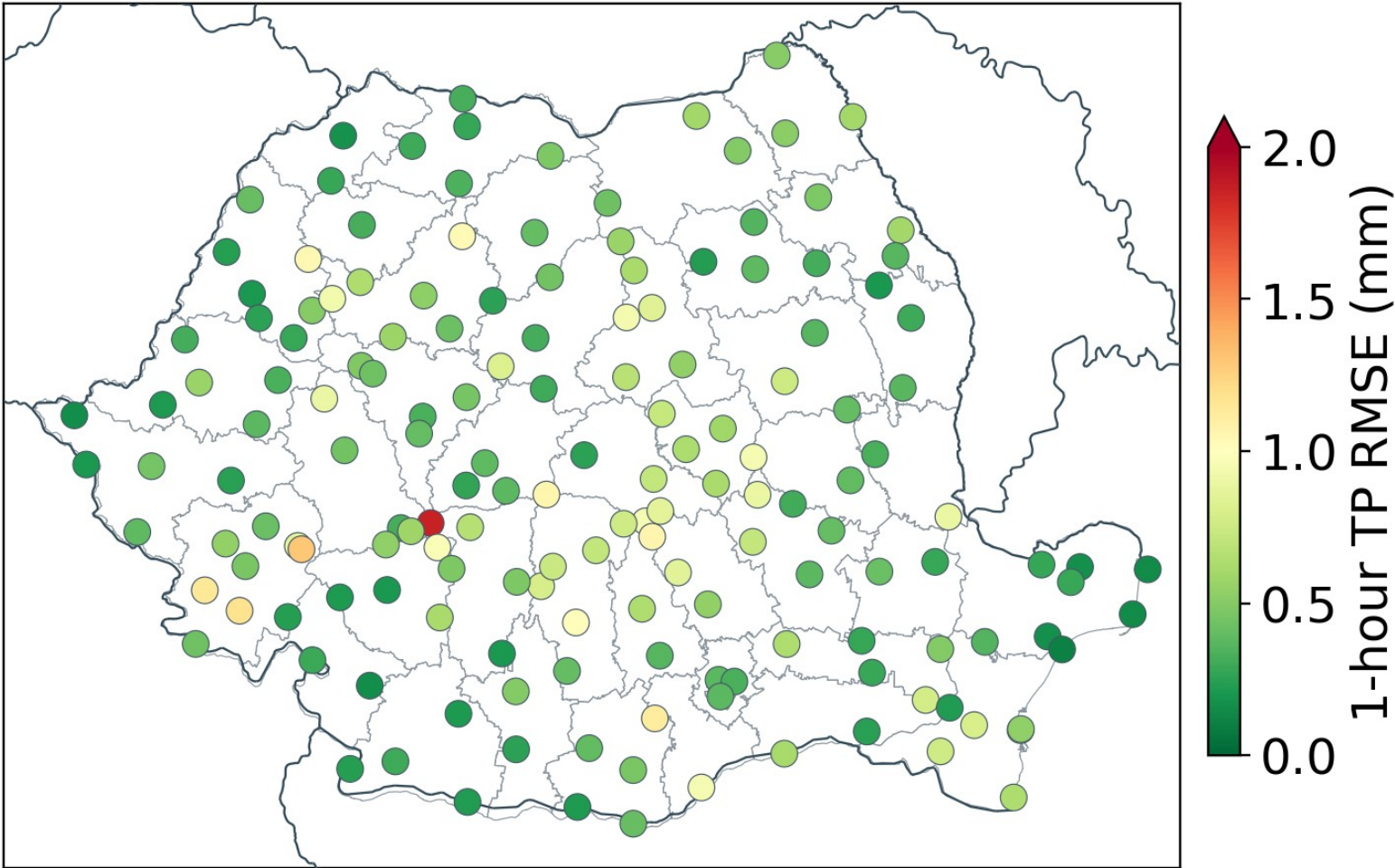
★ **GIURGIU**

Raw 63: moderate continuous rain

Calibrated 81: moderate or heavy rain showers

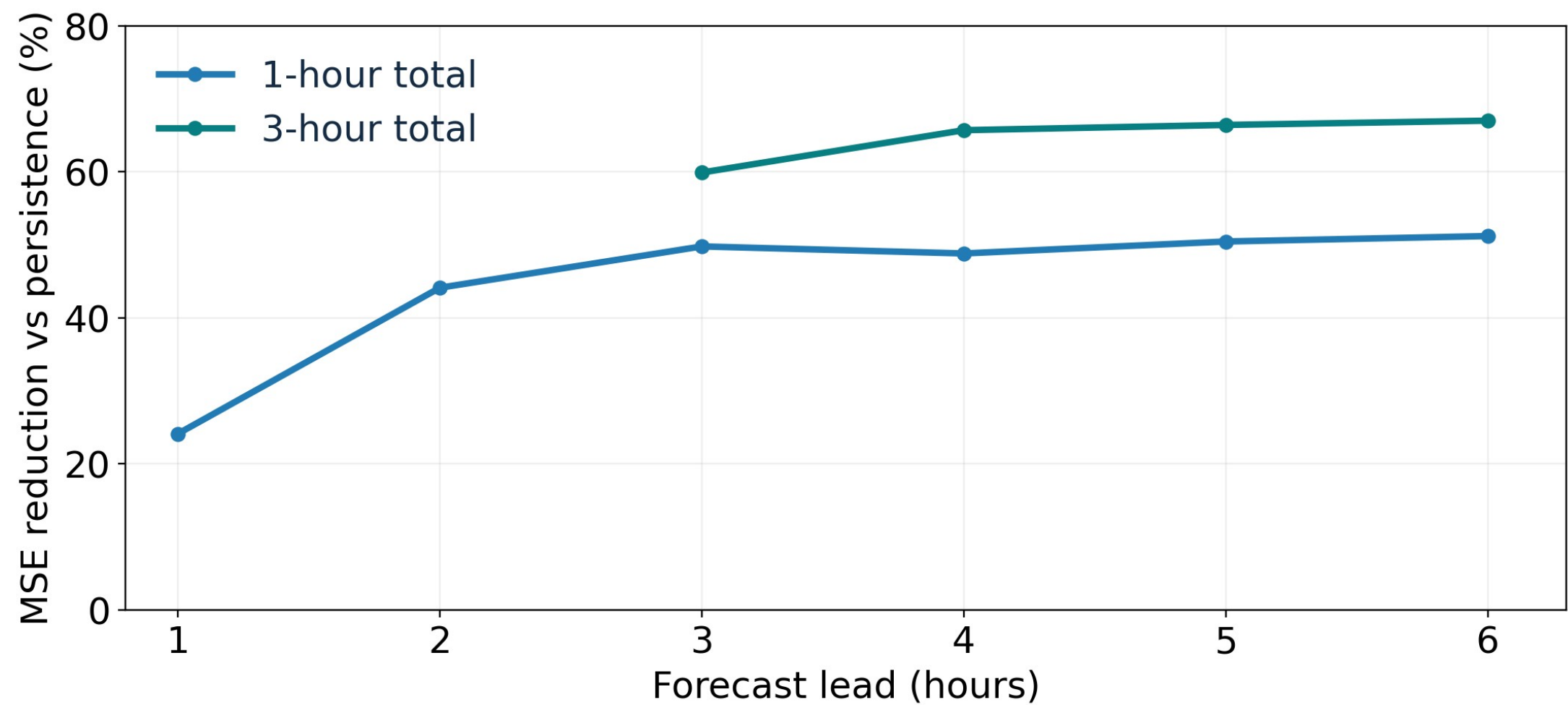
SYNOP 80: slight rain showers

806 initializations verified at 162 rain stations



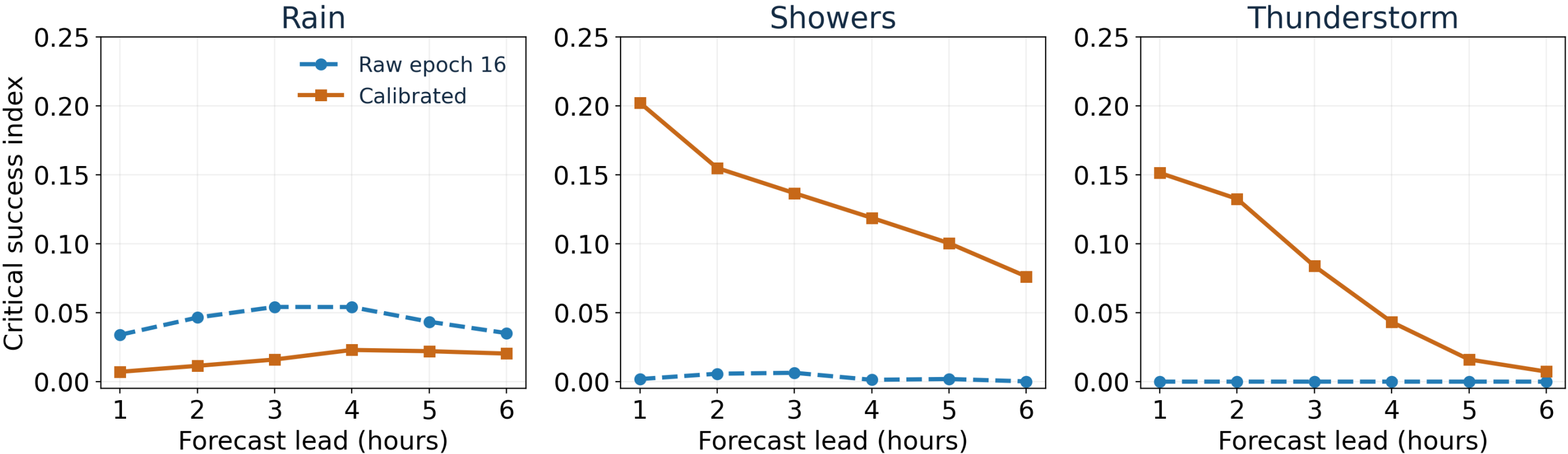
One full-period epoch-16 sample supports the raw-calibrated comparison.

Lower error than persistence does not imply precise storm timing



+6 h: 51% lower MSE for 1 h TP; 67% for 3 h TP, versus persistence.

Calibration expands early shower and thunderstorm detection



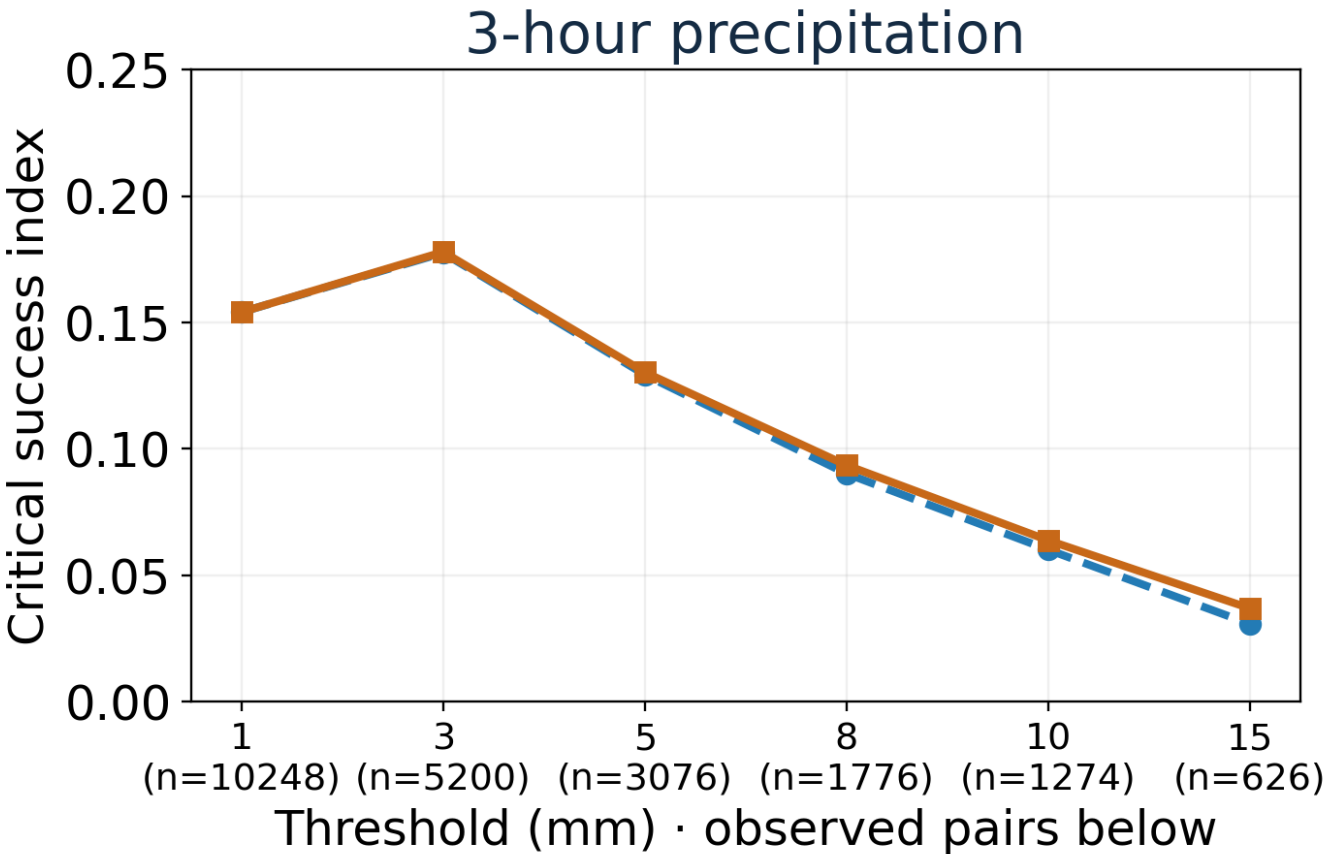
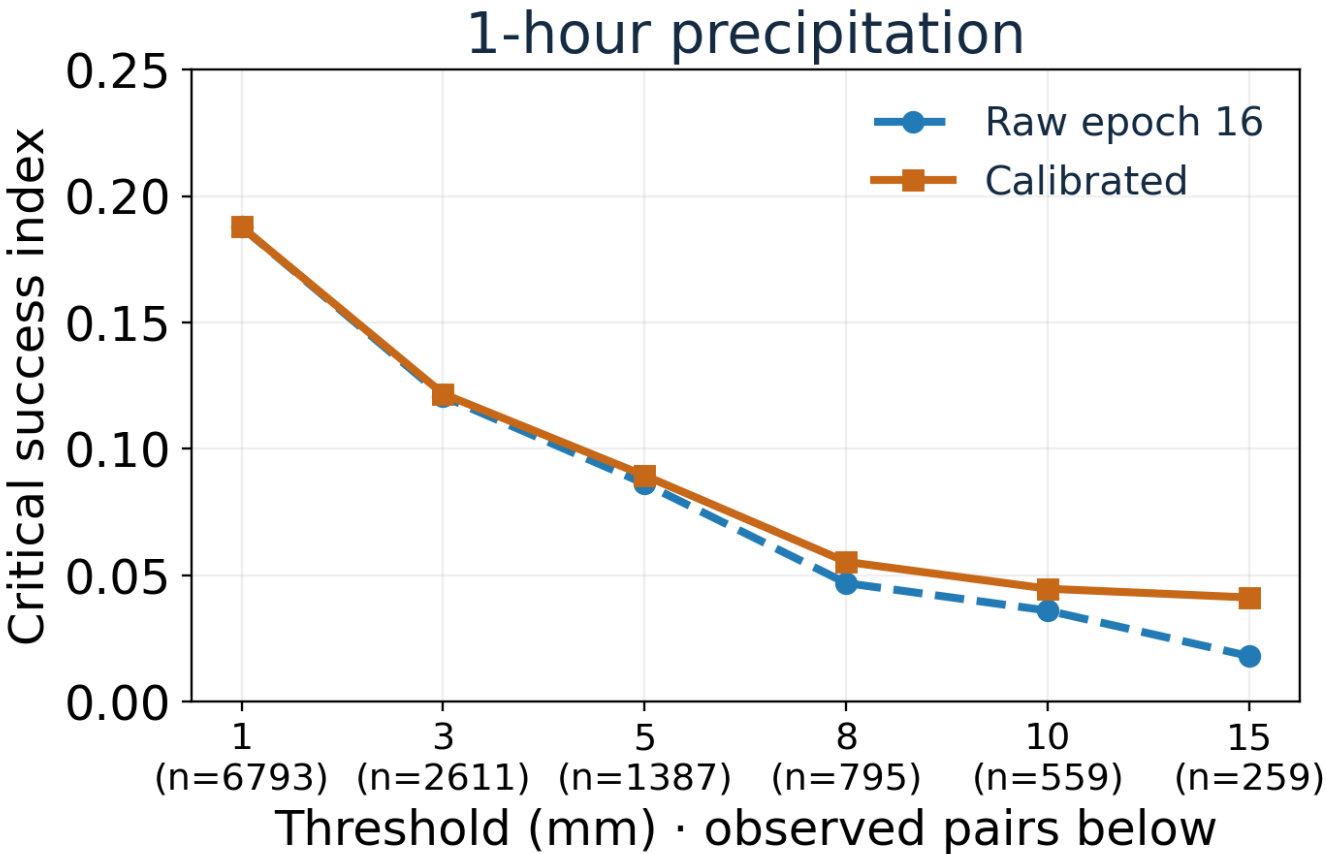
Convective-weather gains are strongest early; rain-category scores mostly decrease.

Three-hour WW verification at the nearest grid cell

Observed weather	Model	POD (%) ↑		FAR (%) ↓		Hits	Misses	Forecasts
No significant weather / cloud n = 112,793	TAU raw		78.0		6.2	88,030	24,763	93,851
	Calibrated		77.9		5.7	87,819	24,974	93,078
	Ensemble		97.8		8.0	110,365	2,428	119,976
Fog n = 3,690	TAU raw		9.8		94.4	363	3,327	6,469
	Calibrated		9.8		94.4	363	3,327	6,460
	Ensemble		0.1		97.6	5	3,685	210
Rain n = 2,136	TAU raw		49.9		95.4	1,065	1,071	23,324
	Calibrated		14.0		98.3	298	1,838	17,486
	Ensemble		17.0		90.6	364	1,772	3,861
Showers n = 3,462	TAU raw		1.2		69.3	42	3,420	137
	Calibrated		46.6		70.8	1,612	1,850	5,527
	Ensemble		0.0		100.0	0	3,462	26
Solid precipitation n = 4	TAU raw		0.0		100.0	0	4	35
	Calibrated		0.0		100.0	0	4	9
	Ensemble		0.0		100.0	0	4	47
Thunderstorms n = 2,019	TAU raw		0.0		N/A	0	2,019	0
	Calibrated		20.0		67.9	403	1,616	1,257
	Ensemble		0.1		57.1	3	2,016	7

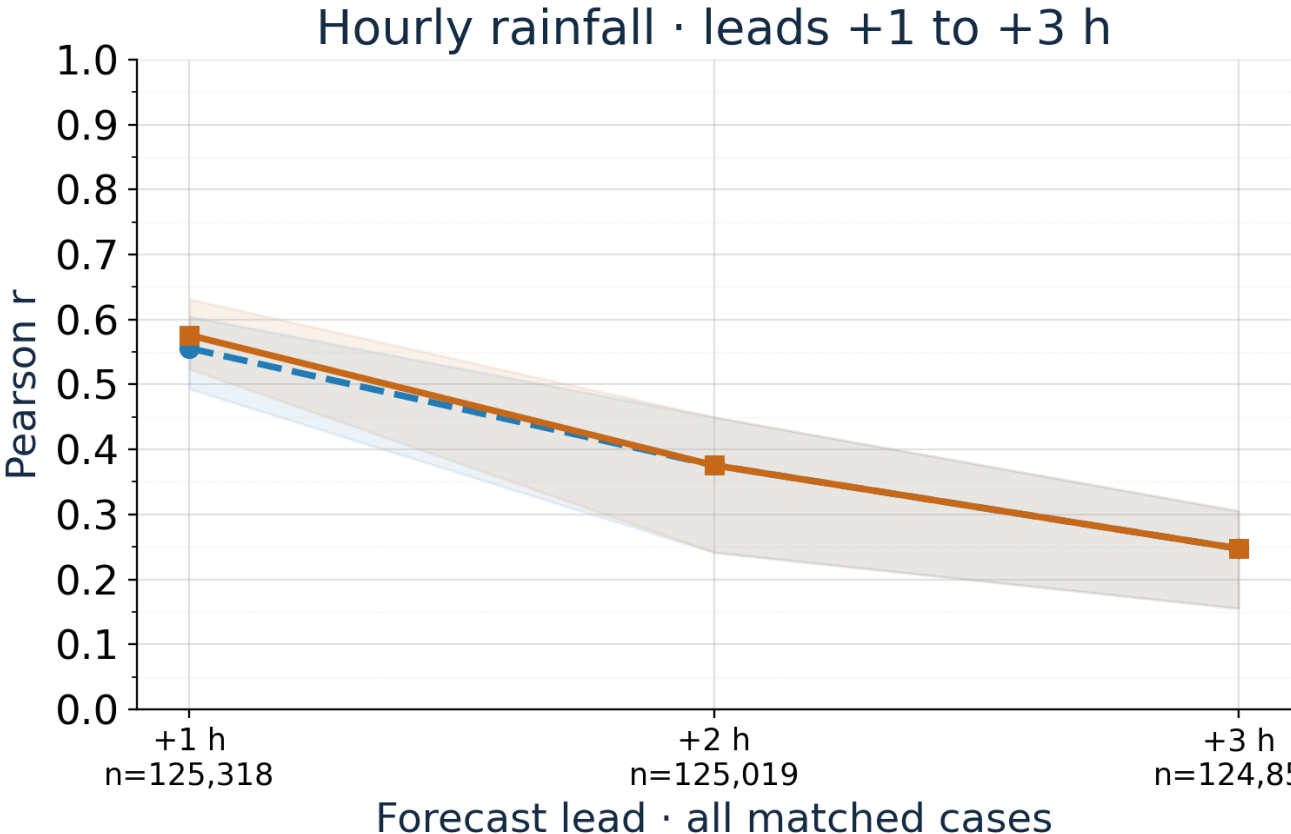
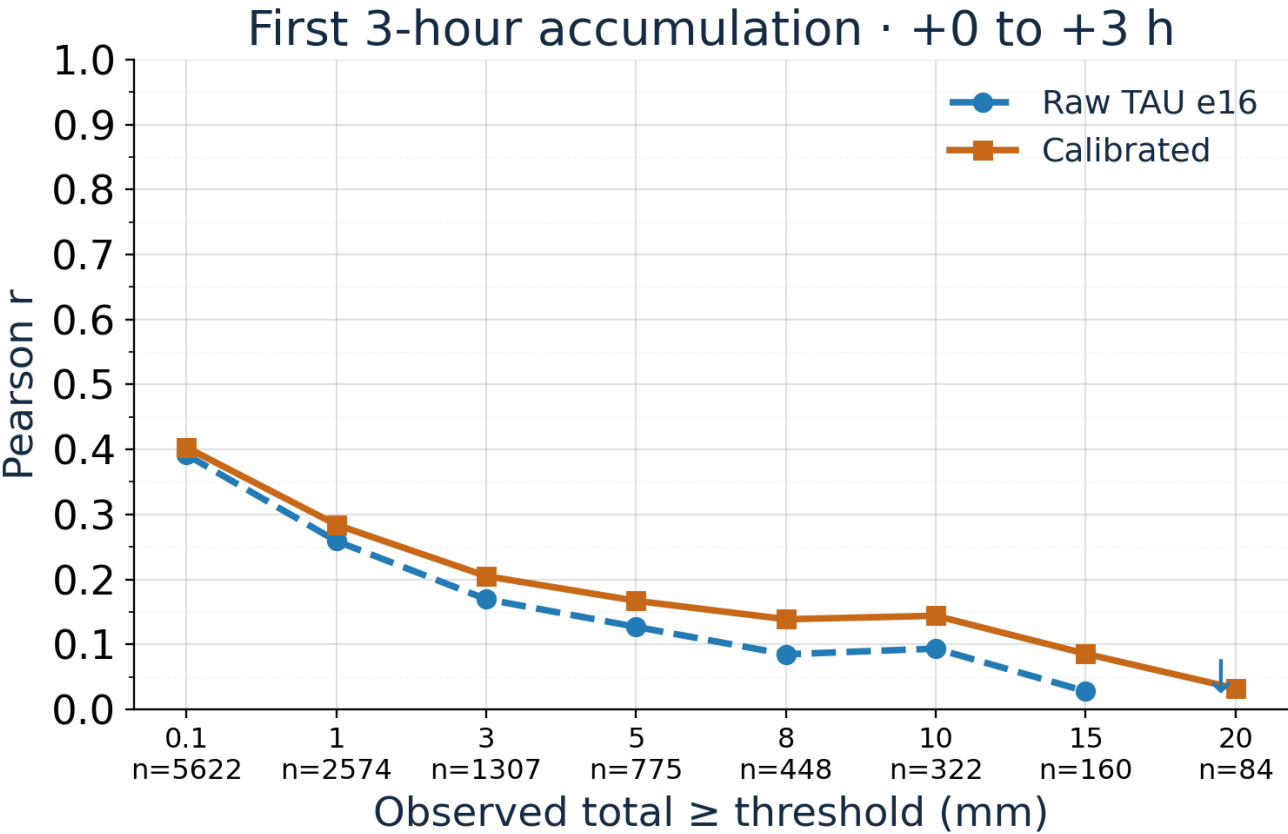
First 3 h · nearest grid cell · TAU epoch 16 · Ensemble: U-Net + two GraphCast-style models
Full archive · retrospective · 800 initializations · 124,127 matched station windows · n = observed windows
Bars: 0–100% · POD: higher is better · FAR: lower is better · N/A: no forecasts

Heavy-rain calibration gains remain limited



Calibrated ≥ 5 mm / 1 h: 12.8% detected • 77.2% false-alarm ratio • CSI 0.089.

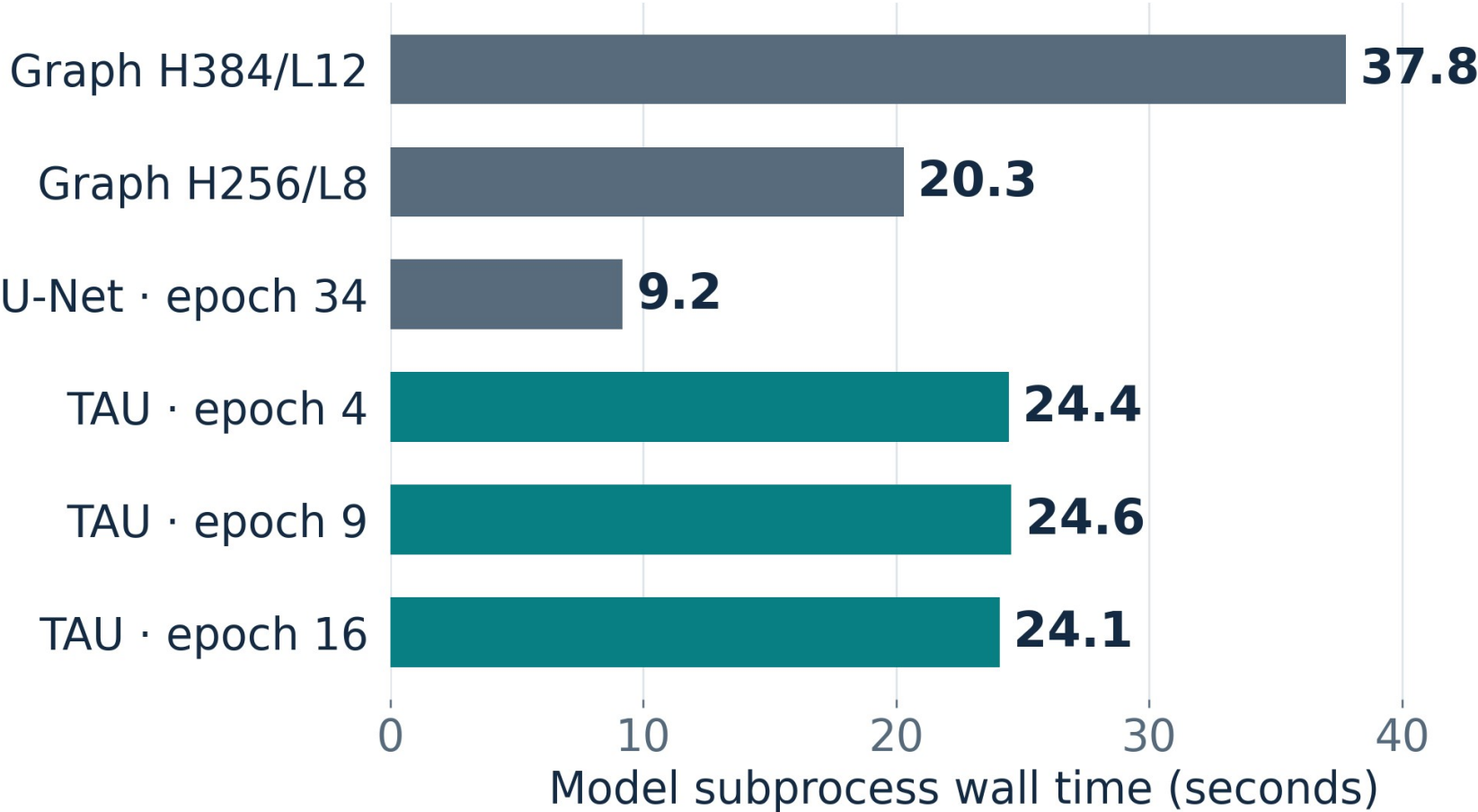
Six-hour output, strongest agreement in the first hours



Full archive · retrospective · n = matched pairs · ↓ = below range · shading: day-block 95% intervals

Prioritise early nowcasting; improving hours 4–6 is the next training target.

Six model products are generated affordably on the CPU



140.4 s

six model
subprocesses

222.6 s

full model /
product stage

Load + forecast calculation + saved arrays + quicklook; additional product generation is separate.

CPU tuning cuts energy per inference by 42%

Same model and historical input • i7-11700 / UHD 750 • FP32

Configuration	Forward (s)	Package (W)	Energy (J)
CPU • 4 threads	15.69	65.0	1,022
CPU • 8 threads + layout tuning	9.05	65.0	595
OpenVINO • CPU/iGPU hybrid	17.13	65.0	1,112
Intel PyTorch • full-model iGPU	12.75	68.7	876

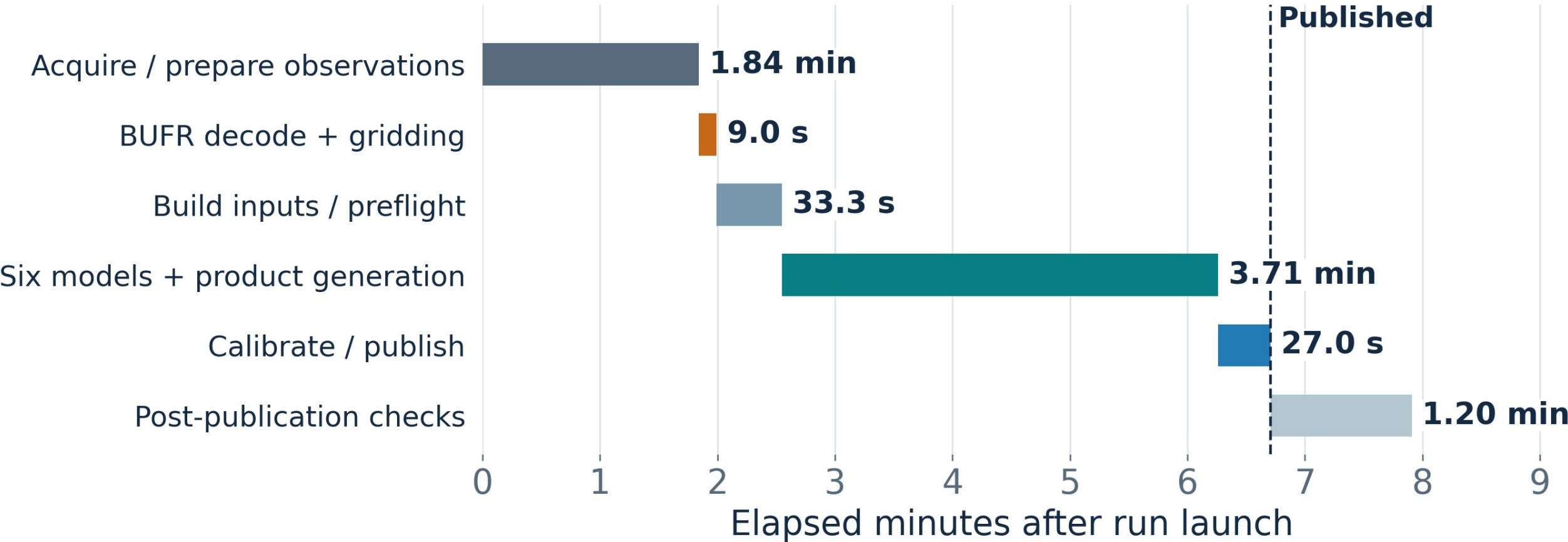
iGPU: 14% less energy than CPU 4 • Optimized CPU: lowest time and energy

Median of 3 warmed forwards per configuration. RAPL package estimates include background activity.
J = joules (energy); W = joules/second (power). Loading and publication are outside this measurement.

Finishing the same forecast sooner reduced energy use, even at similar package power.

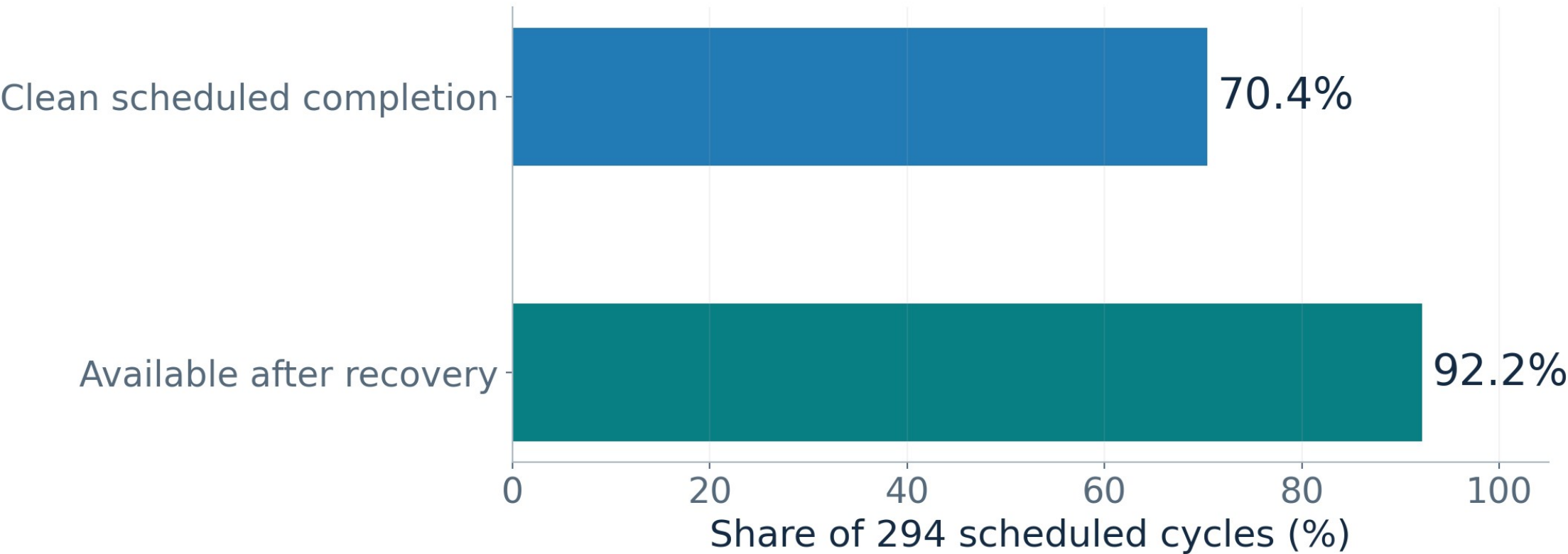
Forecast products reach publication in 6.7 minutes

Launched 14:30:03 UTC → published 14:36:45 UTC



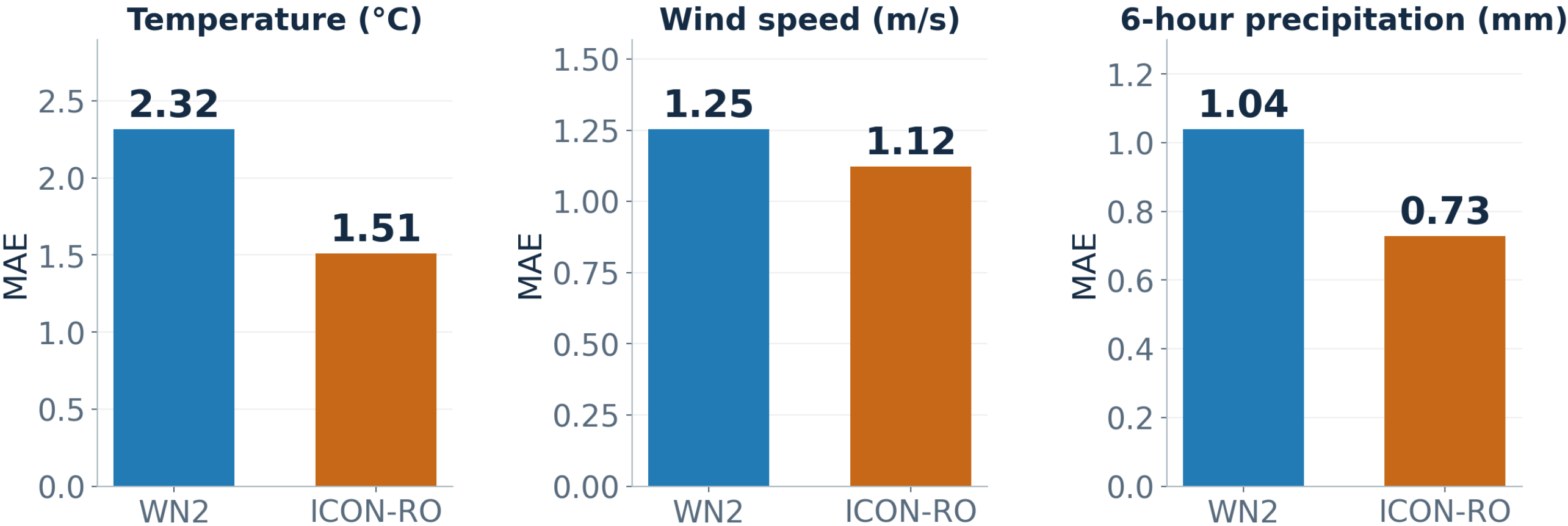
BUFR preparation: 9.0 s • Publication: 6.7 min after launch • Scheduled 30-minute offset is separate.

Recovery is part of the forecasting design



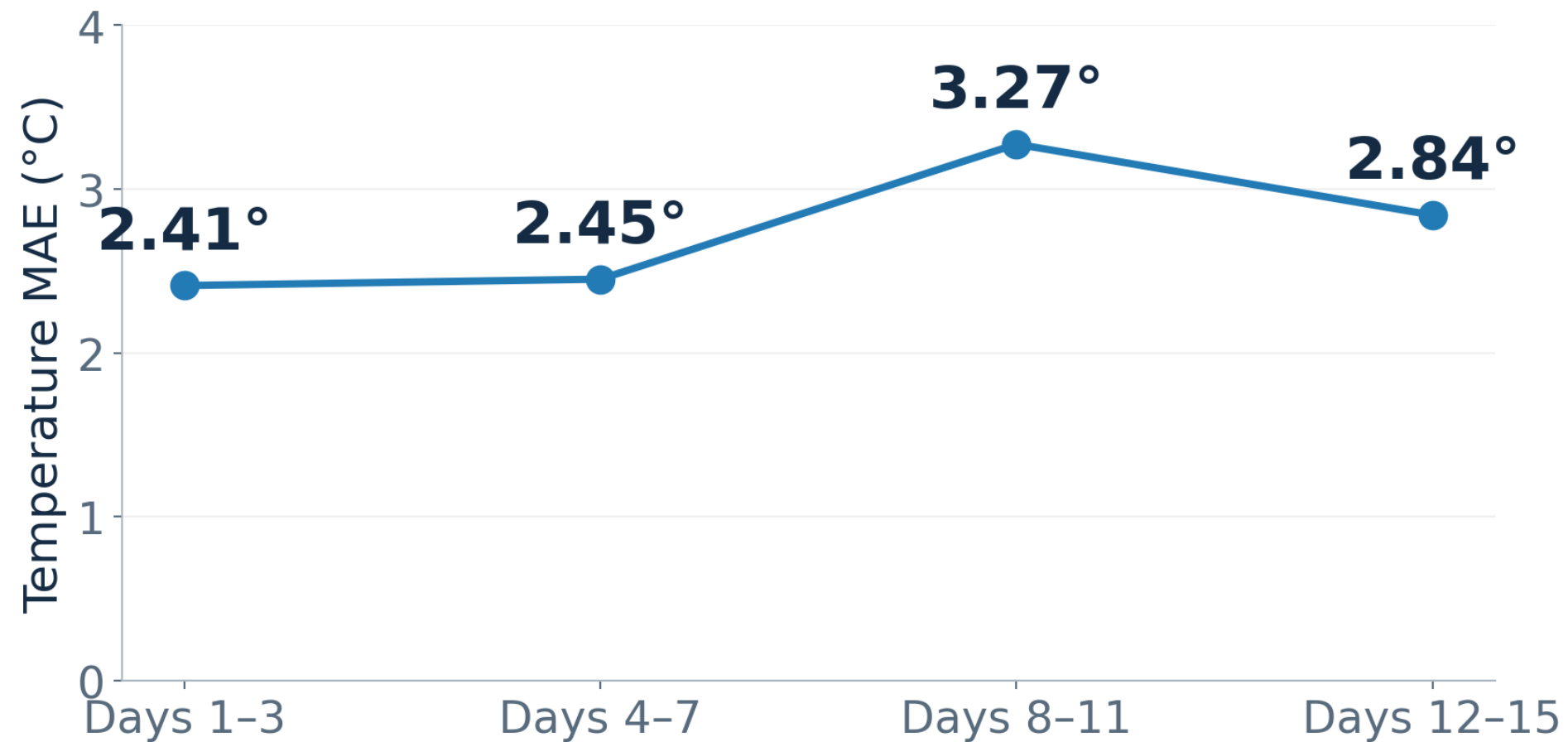
Timeouts, resumable stages and backfill protect continuity; timeliness needs its own score.

ICON-initialized WN2 extends the study to 15 days



The station benchmark identifies temperature, wind and rainfall correction targets against ICON-RO.

Later-day temperature errors are regime-dependent



Fixed 144-station network • one complete 21 August trajectory • error is not monotonic with lead.

Full-resolution WN2 is practical on local hardware

CPU • 48 h trajectory

106.3 min mean rollout
Two completed runs
FP32 • CPU / XLA

2 GPUs • 360 h trajectory

11 min 35 s rollout + save
14 min 54 s full process
BF16 • cached executable

Corrected GPU footprint

29.33 GiB host RAM
16.24 / 15.53 GiB VRAM
11.04 s warm step

Forecast execution, startup / compilation, output writing and total process time are separate measurements.

Extend early-hour nowcasting skill toward hours 4–6

Longer-lead storm skill

- Improve hours 4–6
- Retain intensity and location
- Test false alarms explicitly

Independent selection

- Prospective event blocks
- Lead-wise calibration gates
- Training / live-input parity

Operational value

- Publication deadline scores
- ICON short-forecast trials
- Corrected WN2 verification

Six-hour output is the design horizon; reliable storm guidance must be demonstrated at each lead.

WINDMIL-AI is now a measurable national AI capability

National observations → early-hour nowcasting → measured forecast performance

Already established

Hourly pre-production

Strongest agreement in early hours

Reusable NWP / global-AI interfaces

Next operational milestone

Improve hours 4–6

Better storm detection

Reliable publication deadlines

Thank you for your attention

Turn inference savings into operational evidence

Established here

12 measured forwards completed
No GPU faults recorded
CPU tuning: 42% less energy
iGPU vs CPU 4: 14% less energy

Next operational test

Replay the tuned CPU candidate
Check outputs and delivery time
Measure the full cycle separately
Test iGPU concurrency explicitly

Training: test these settings on forward + backward + optimizer steps;
inference savings do not establish GPU training speed or epoch duration.

One historical case • fixed-order pilot • package energy includes background activity
Background after experiments: 52.4 W package; no idle subtraction or system changes.

Promote the measured CPU gains through an isolated replay, then verify whole-workflow value.