

Model Verification at ARP AE

01

Status of FFV2-based
verification at ARP AE

Deterministic and ensemble
systems

Authors

Thomas Gastaldo, Davide Cesari,
Alessandro Frigato, Maria Stefania
Tesini

02

Spatial verification of
precipitation

Methods and applications using
FSS and User-Oriented
Approaches

Authors

Leonardo Beltrami, Maria Stefania Tesini

Model Verification at ARPAE

01

Status of FFV2-based
verification at ARPAE

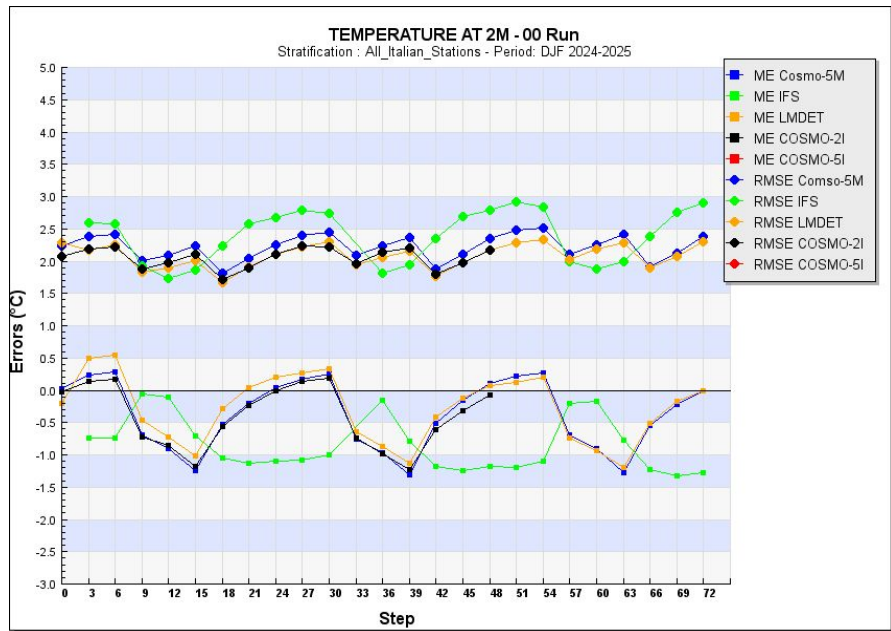
Deterministic and ensemble
systems

Authors

Thomas Gastaldo, Davide Cesari,
Alessandro Frigato, Maria Stefania
Tesini

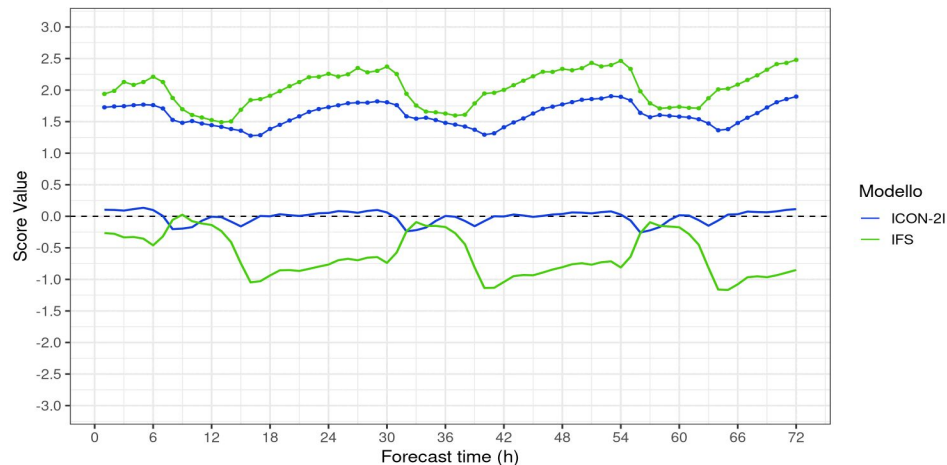
From Versus to FFV2

- **Surface verification was temporarily suspended during the transition to ICON and resumed with FFV2 in 2025.**
- No direct COSMO–ICON verification is available; however, comparable station datasets were used to assess consistency between the two verification systems.
- **Example: T2m verification for DJF 2024–25 with VERSUS and FFV2.**



DJF_2024-25: ICON-2I vs IFS – stat_all-ita – T2M (°C) - ME/RMSE

ME: linea continua RMSE: linea con punti



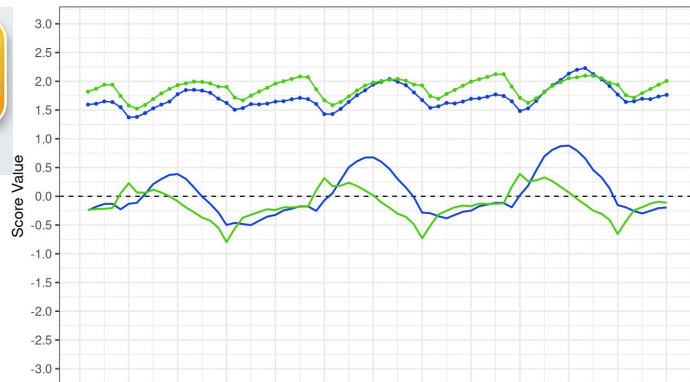
Example: Seasonal Verification of 2m Temperature



JJA

JJA_2025: ICON-2I vs IFS – stat_all-ita – T2M (°C) - ME/RMSE

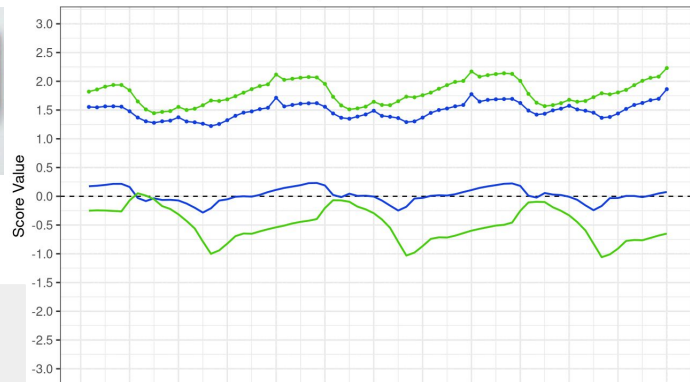
ME: linea continua RMSE: linea con punti



SON

SON_2025: ICON-2I vs IFS – stat_all-ita – T2M (°C) - ME/RMSE

ME: linea continua RMSE: linea con punti



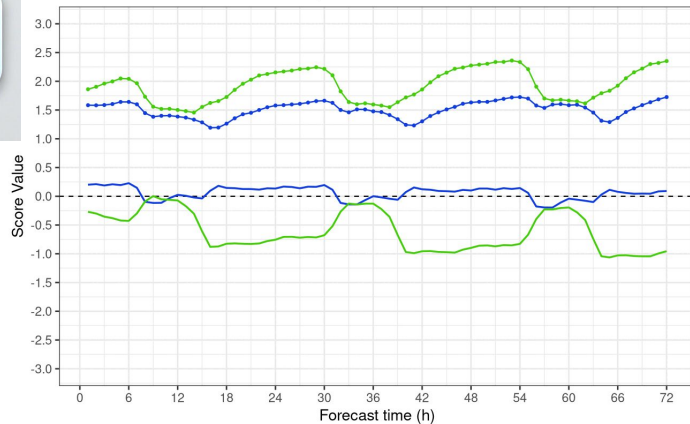
IFS-ECMWF
ICON-2I



DJF

DJF_2025-26: ICON-2I vs IFS – stat_all-ita – T2M (°C) - ME/RMSE

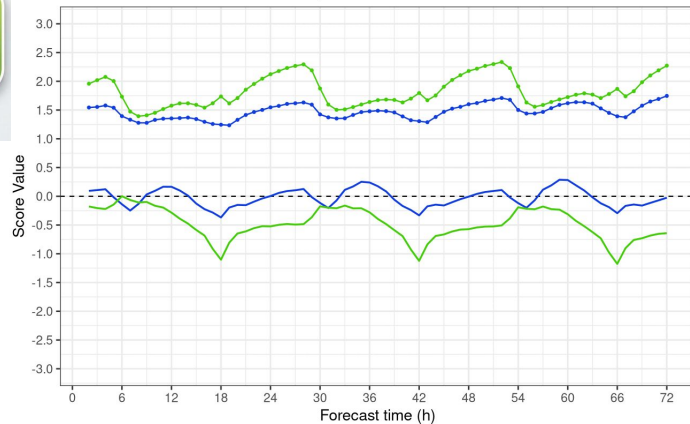
ME: linea continua RMSE: linea con punti



MAM

MAM_2026: ICON-2I vs IFS – stat_all-ita – T2M (°C) - ME/RMSE

ME: linea continua RMSE: linea con punti

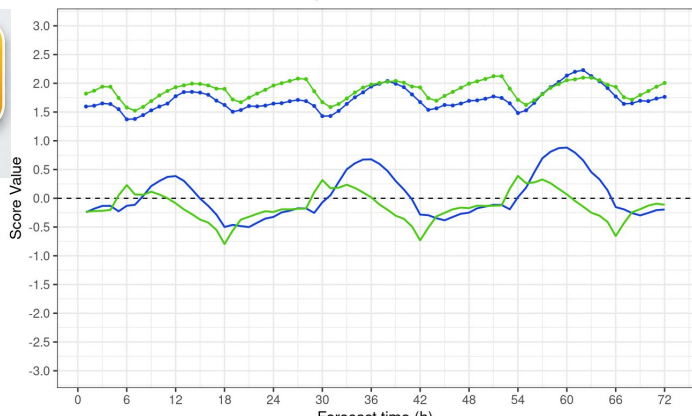


Summer Verification over the Po Valley: Some Issues



JJA
2025

JJA_2025: ICON-2I vs IFS – stat_all-ita – T2M (°C) - ME/RMSE
ME: linea continua RMSE: linea con punti



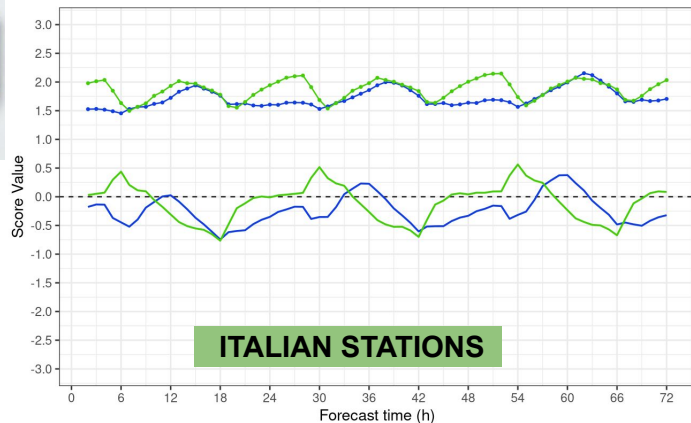
Modello
— ICON-2I
— IFS

IFS-ECMWF
ICON-2I



JJA
2026

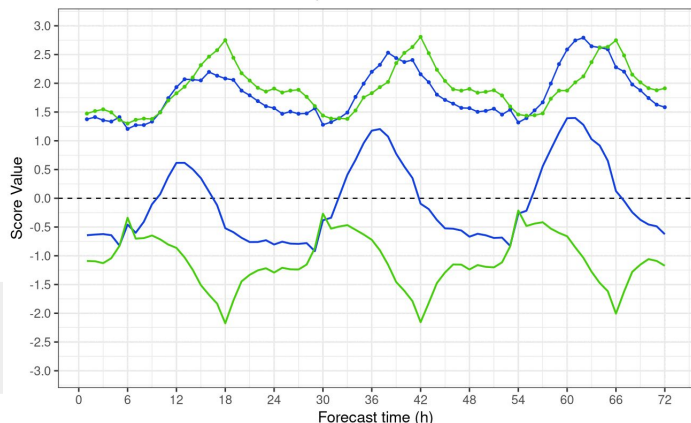
JJA_2026: ICON-2I vs IFS – stat_all-ita – T2M (°C) - ME/RMSE
ME: linea continua RMSE: linea con punti



Modello
— ICON-2I
— IFS

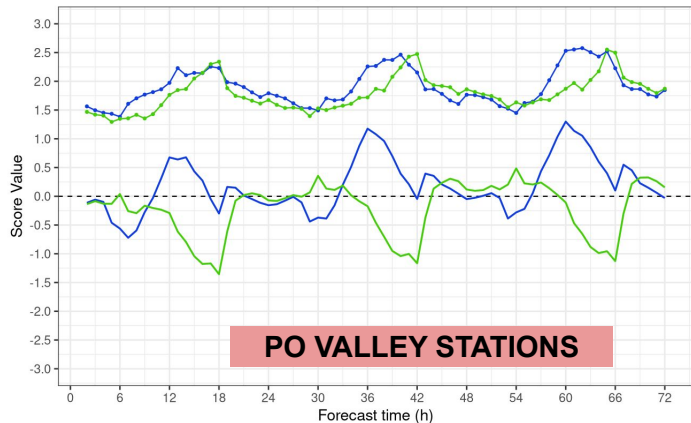
ITALIAN STATIONS

JJA_2025: ICON-2I vs IFS – stat_poavalle – T2M (°C) - ME/RMSE
ME: linea continua RMSE: linea con punti



Modello
— ICON-2I
— IFS

JJA_2026: ICON-2I vs IFS – stat_poavalle – T2M (°C) - ME/RMSE
ME: linea continua RMSE: linea con punti

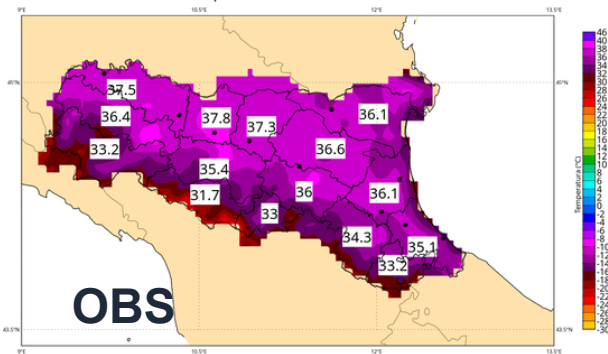


Modello
— ICON-2I
— IFS

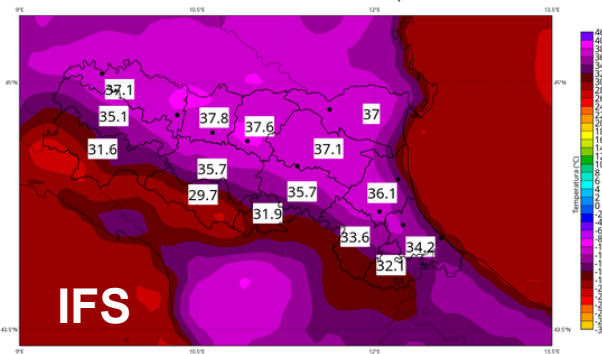
PO VALLEY STATIONS

Impact of T2m Overestimation on Heat Warnings

TMax giornaliera (°C) da interpolazione ERG-GIAS + 50° percentile
Osservazione valida per il 27-06-2026 dalle 00 alle 24 UTC 2m

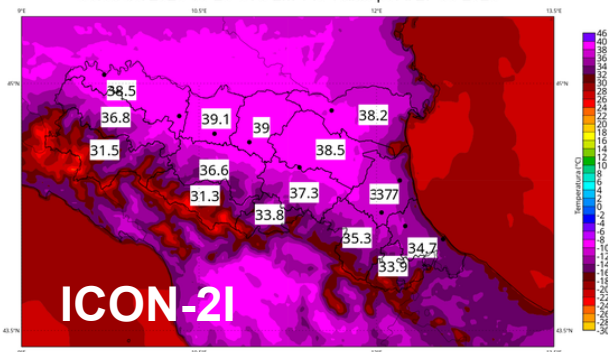


TMax giornaliera prevista da ECMWF (°C) + 50° percentile
Corsa del 2026-06-26 00:00 UTC 2m +48 valida per il 27-06-2026

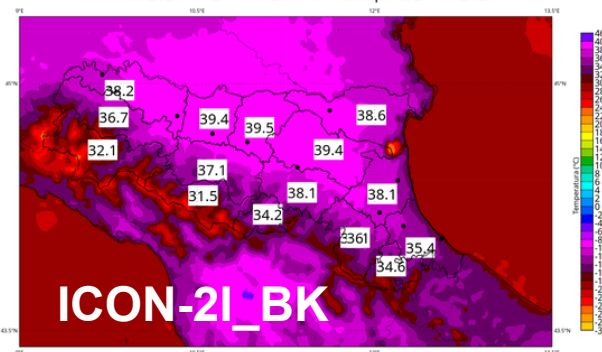


Tmax overestimation was frequently observed during last summer's hot spells, especially at longer lead times

TMax giornaliera prevista da ICON 2I (°C) + 50° percentile
Corsa del 2026-06-26 UTC 2m +48 valida per il 27-06-2026



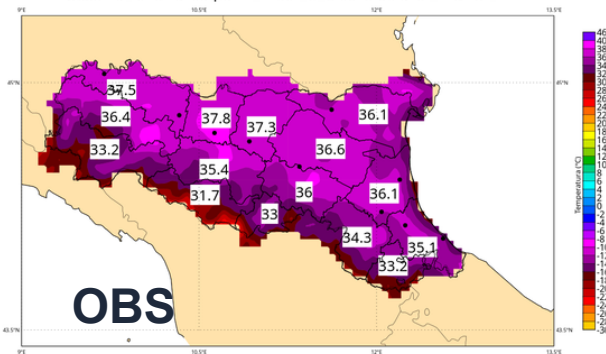
TMax giornaliera prevista da ICON 2I Backup (°C) + 50° percentile
Corsa del 2026-06-26 UTC 2m +48 valida per il 27-06-2026



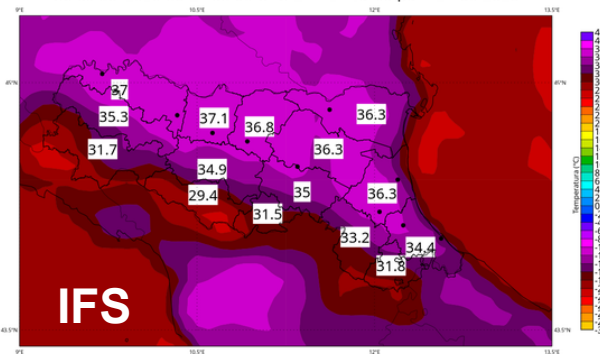
Valid for 27 June 2026
*Forecast issued on
26 June — Day 2*

Impact of T2m Overestimation on Heat Warnings

TMax giornaliera (°C) da interpolazione ERG-GIAS + 50° percentile
Osservazione valida per il 27-06-2026 dalle 00 alle 24 UTC 2m

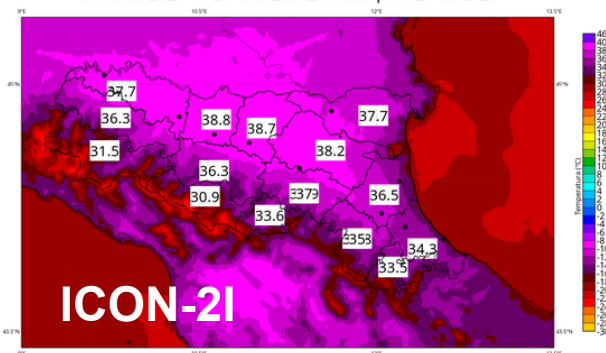


TMax giornaliera prevista da ECMWF (°C) + 50° percentile
Corsa del 2026-06-27 00:00 UTC 2m +24 valida per il 27-06-2026

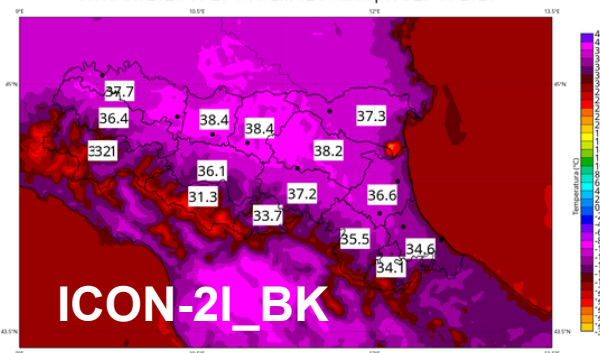


During last summer's extreme heat events, Tmax overestimation appeared to increase at the highest temperatures. This is one of the aspects we plan to investigate further in the coming months using conditional verification

TMax giornaliera prevista da ICON 2I (°C) + 50° percentile
Corsa del 2026-06-27 UTC 2m +24 valida per il 27-06-2026



TMax giornaliera prevista da ICON 2I Backup (°C) + 50° percentile
Corsa del 2026-06-27 UTC 2m +24 valida per il 27-06-2026



Valid for 27 June 2026
*Forecast issued on
27 June — Day 1*

Model Verification at ARPAE

01

Status of FFV2-based
verification at ARPAE

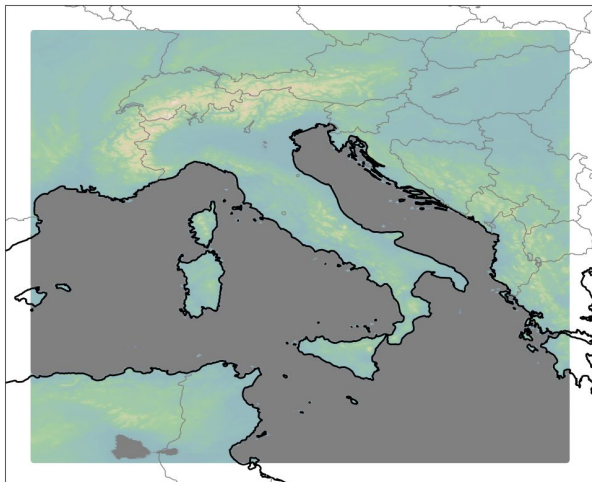
Ensemble systems

Authors

Thomas Gastaldo, Davide Cesari,
Alessandro Frigato, Maria Stefania Tesini

ICON-2I-EPS : Summer verification against IFS-ENS

ICON-2I-EPS Setup

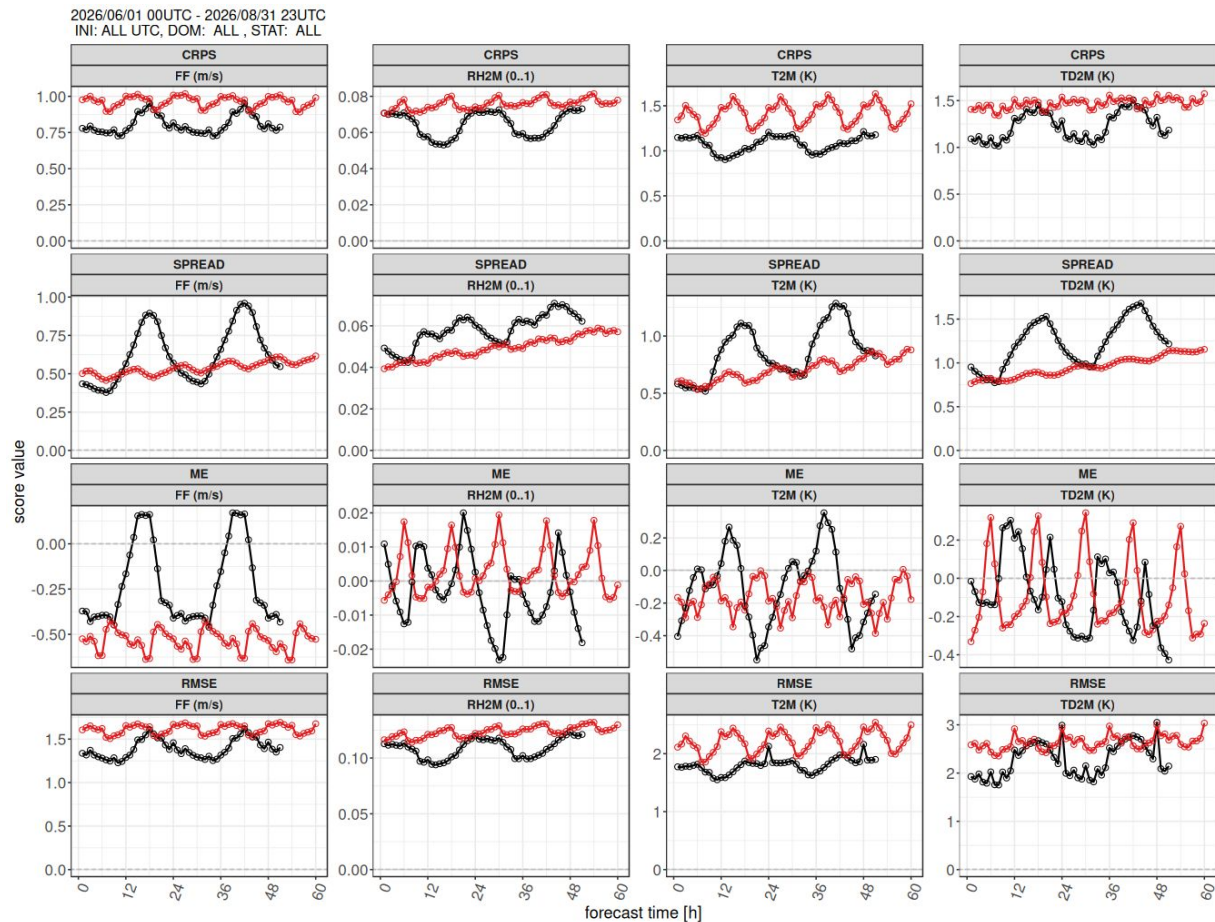


- Grid horizontal res. : 2.2 km
- v.l : 65
- I.Cs: KENDA
- BCs : from IFS' first 20 members
- Init. Time : 21 UTC
- Time range: 51h
- Verification is not aligned for iniTimes (3h shift)

Namelist

```
# SYNOP eps namelist (icon_2I - ifs settings)
varnoContinuous 'T2M,RH,RH2M,PS,TD2M,FF,DD,RAD_GL_1h,RAD_DF_1h,TMIN_12h,TMAX_12h,JJ_12h,N,N_L,N_M,N_H'
pecthresholds list('PS'=list('lower'=c(-500),'upper'=c(500)), 'FF'=list('lower'=c(-3),'upper'=c(3)))
catthresholds list('GUST_1h'=c(5,10,15), 'RR_1h'=c(0.1,1,2), 'T2M'=c(290,300,310))
subdomains 'ALL'
filePattern 'vepSYNOP.'
timeSteps '0'
timeBreaks '0,-60'
veri_run_type '0,4,3'
veri_run_class '0,2'
state '0,1,5,11'
#r_state '0,1,5,11'
sigTest 'T'
expType 'standard'
experiment 'T'
expIds 'ICON-2I,IFS'
expDescr '2026jja_I2I_IFS'
alignObs 'F'
fileDescr '2026jja_I2I_IFS'
fdbkDirs '/autofs/scratch-mod/afrigato/verifica_FF/icon_2I-eps/2026-jja,/autofs/scratch-mod/afrigato/verifica_FF/ifs-ens/2026-jja'
outDir './'
shinyAppPath '/autofs/nfshomes/afrigato/ShinyApps'
nMembers '20'
```

2026 JJA : Lead Time



IniTimes:
ICON-2I: 21 UTC
IFS: 00, 12 UTC

Sig. diff. (95%)

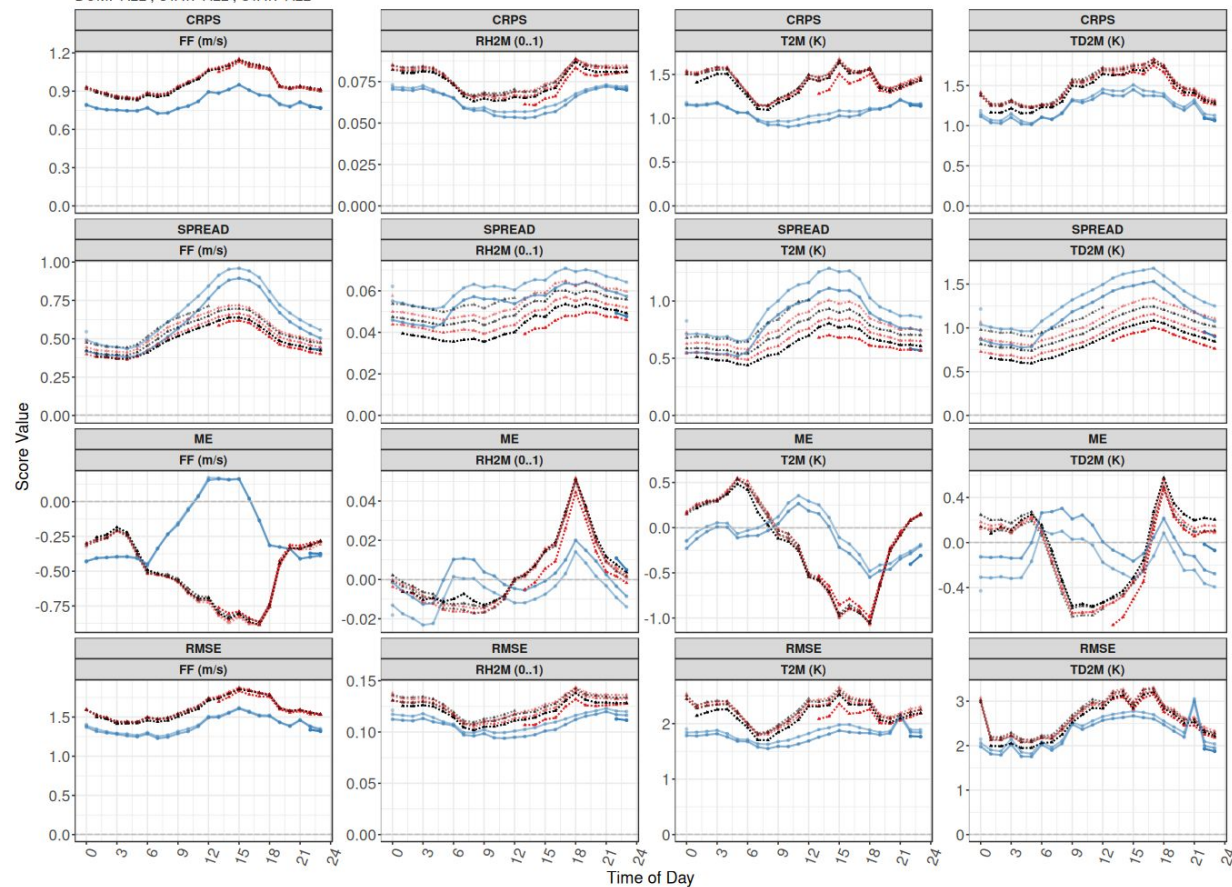
○ n.a.
no
yes

Experiment

○ ICON-2I
○ IFS

2026 JJA : DayTime

2026/06/01 00UTC - 2026/08/31 23UTC
DOM: ALL, STAT: ALL, STAT: ALL



Experiment

— ICON-21

- - - IFS

Init

— 00

— 12

— 21

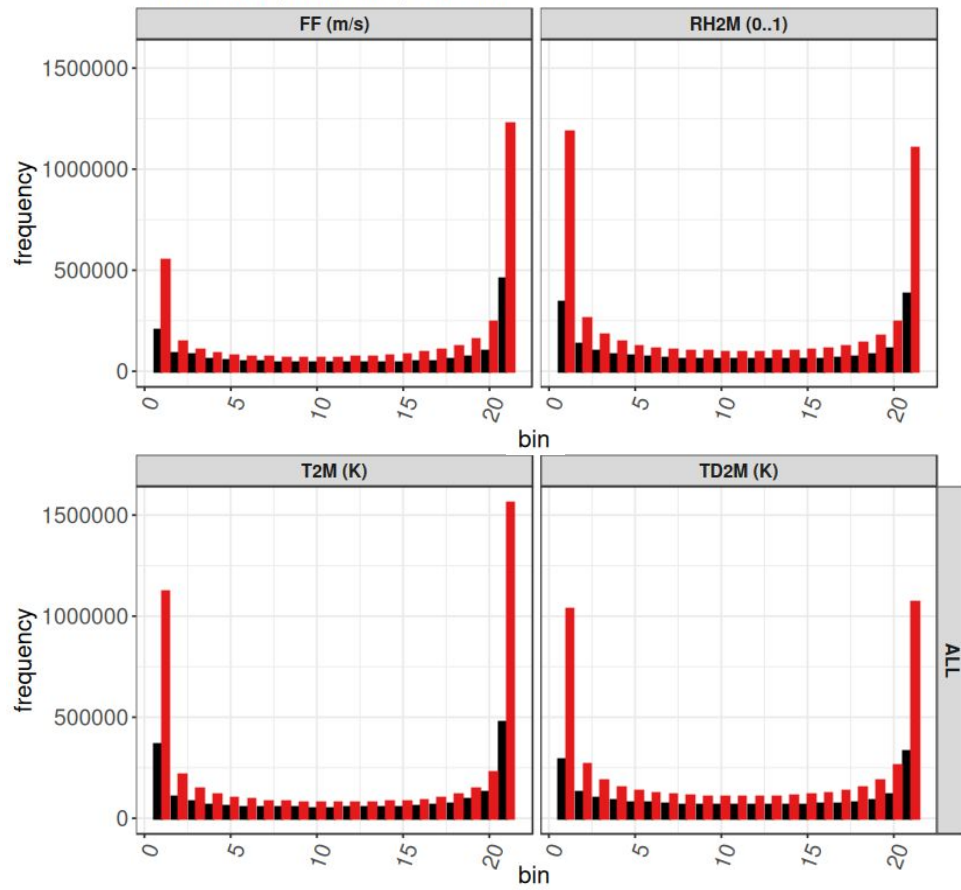
IniTimes:

IFS: 00 12 UTC

ICON-21: 21
UTC

2026 JJA : rank histogram

2026/06/01 00UTC - 2026/08/31 23UTC
INI: ALL UTC, DOM: ALL, STAT: ALL



- Operational:
 - Verification and spatial verification to better represent model uncertainty in extreme weather, convection
 - Implement a verification setup with same lead Times to compare ICON-2I to IFS
- Research [PP KAOS]
 - Task 3.1 : Cosmo software to verify ICON-LEPS, to be implemented
 - Task 3.2 : Spatial verification and diagnostics for EPSs, specific focus on precipitation

Both points overlap with WGs VA, EPS, PHY

Model Verification at ARPAE

02

Spatial verification of
precipitation

Methods and applications using
FSS and User-Oriented
Approaches

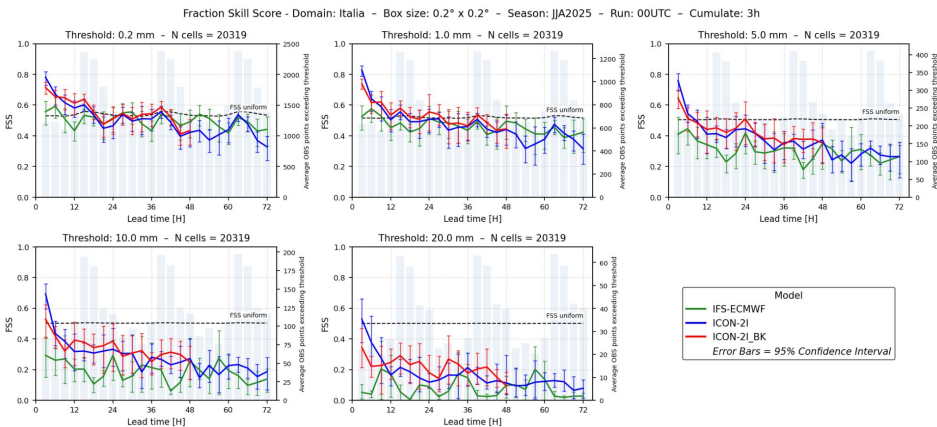
Authors

Leonardo Beltrami, Maria Stefania Tesini

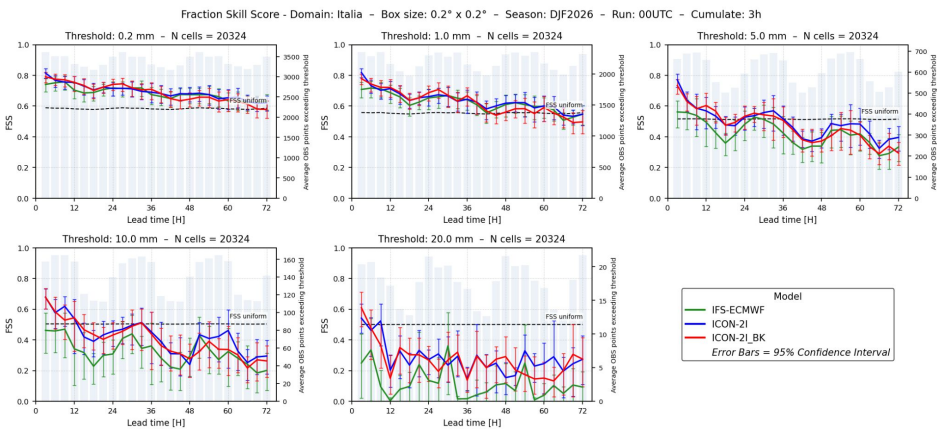
Applications of the Fractional Skill Score to Deterministic and Ensemble Forecasts

Seasonal FSS over the Italian domain

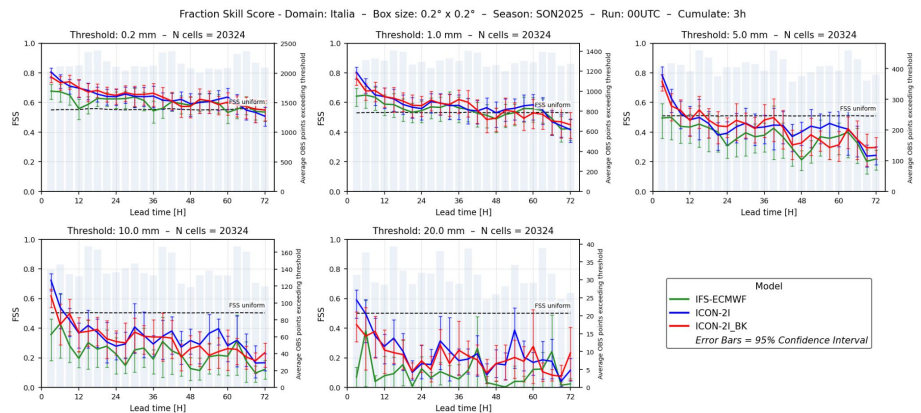
JJA2025



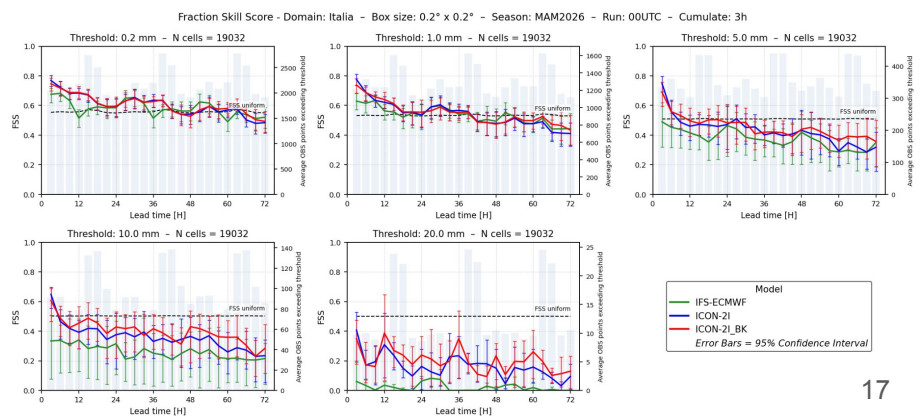
DJF2026



SON2025



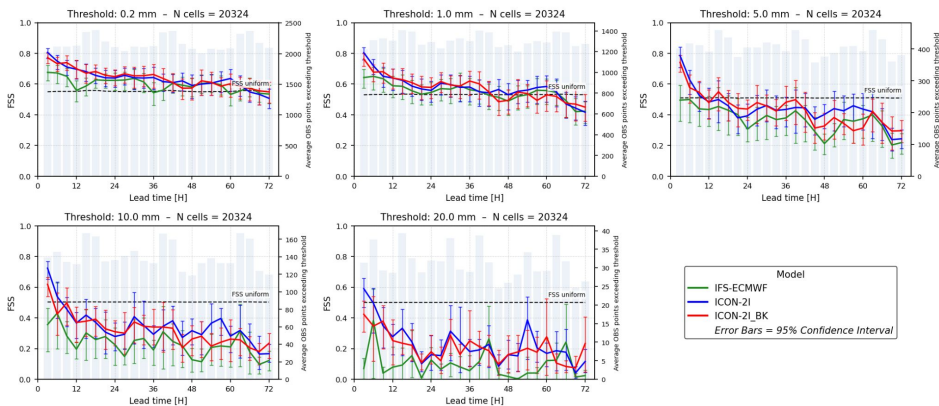
MAM2026



Same season, different regional skill

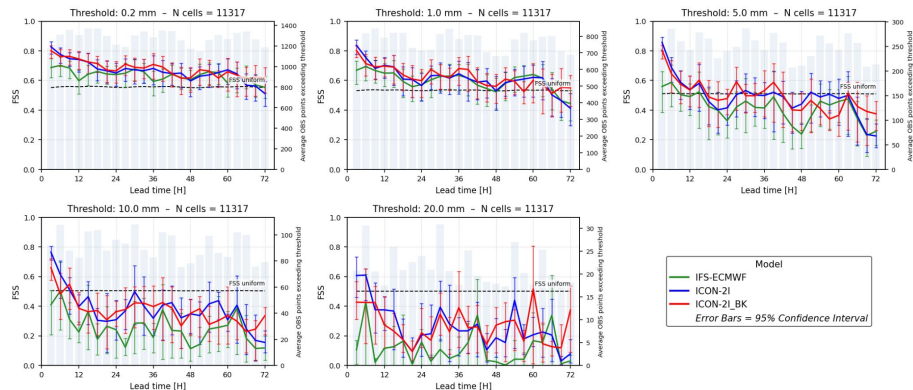
Domain: Italy

Fraction Skill Score - Domain: Italy - Box size: $0.2^\circ \times 0.2^\circ$ - Season: SON2025 - Run: 00UTC - Cumulate: 3h



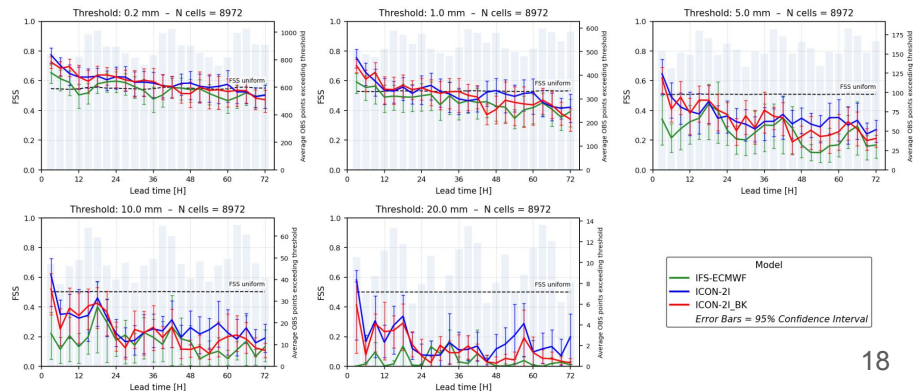
Domain: Northern Italy

Fraction Skill Score - Domain: Nord Italia - Box size: $0.2^\circ \times 0.2^\circ$ - Season: SON2025 - Run: 00UTC - Cumulate: 3h



Domain: Southern Italy

Fraction Skill Score - Domain: Sud Italia - Box size: $0.2^\circ \times 0.2^\circ$ - Season: SON2025 - Run: 00UTC - Cumulate: 3h



1. FSS Skill vs. FSS Spread (Dey et al., 2014)

Skill: every ensemble forecast member is compared against the observation field

Spread: every possible pair of ensemble members is compared against each other

0	$\leq \text{FSS} \leq$	1
no skill max spread		perfect skill no spread

2. Implementation of FSS Skill (Necker et al., 2024, Eq. 7-8)

$$\text{FSS} = 1 - \frac{\sum_{i=1}^I (\text{NEP}_{i,f} - \text{NP}_{i,o})^2}{\sum_{i=1}^I \text{NEP}_{i,f}^2 + \sum_{i=1}^I \text{NP}_{i,o}^2}$$

Where:

f is the forecast (ensemble members) field

o is the observation field (radar corrected with rain gauges)

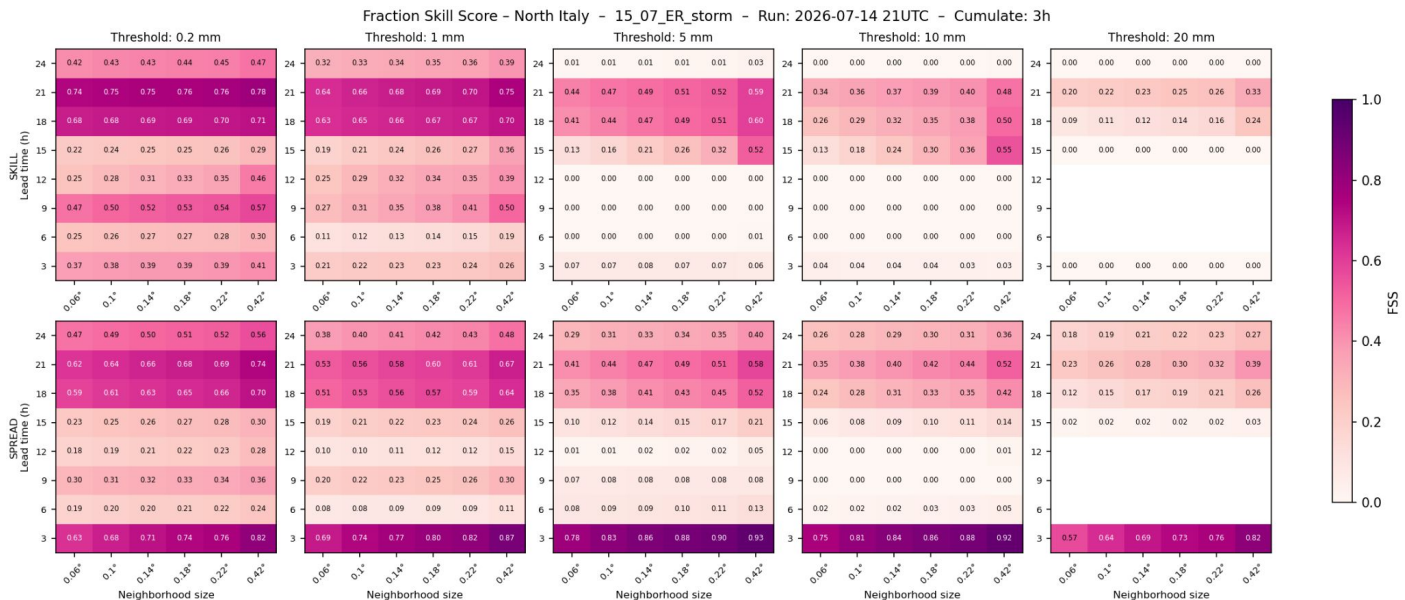
i is the index of a specific instant (i.e. fcst-obs pair)

$\text{NEP}_{i,f}$ is the neighborhood ensemble probability

$\text{NP}_{i,o}$ is the neighborhood probability in the OBS field

3. FSS is implemented using the ‘scores’ python package (Leeuwenburg et al., 2024)

Forecast: Run 14 July 2026 H 21 UTC



FSS is undefined in “dry” cases (Mittermaier, 2021)

If no point exceeds the threshold anywhere, in both the forecast (i.e. all ENS members) and the observation, numerator and denominator vanish together: the FSS is undefined

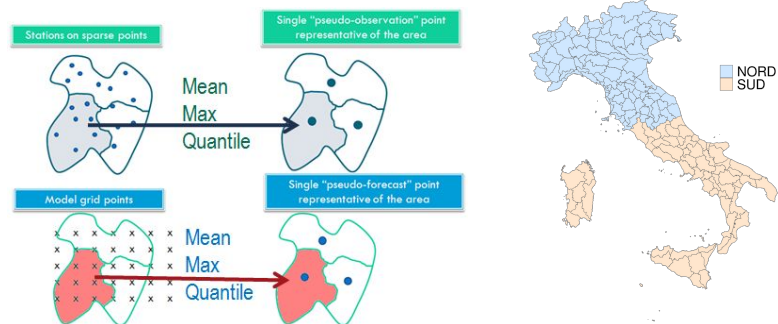


What if the observation field is dry everywhere but, only 1 of the 20 ENS members predicts rain in only 1 point of the domain? The resulting FSS is 0. So: that noise can penalize the metric.

Some minimum threshold could be implemented to cut out noisy cases.

User oriented verification

The verification is performed at the scale of Civil Protection warning areas, using representative values derived from observations and model grid points.

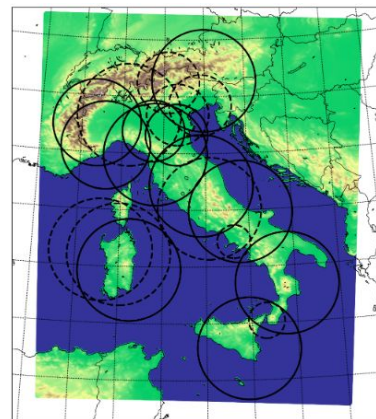


Scores are computed for Italy as a whole and separately for Northern and Southern Italy, based on Civil Protection warning areas.

This allows model performance to be assessed under different geographical and meteorological regimes.

Dataset

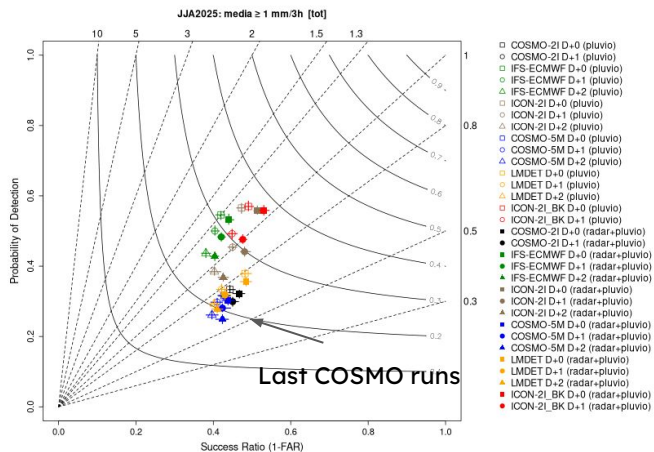
- **Regional & National Networks from DPCN:** High-resolution rain-gauge networks
- **Radar Composite:** Gauge-adjusted national radar composite for 3-h precipitation.



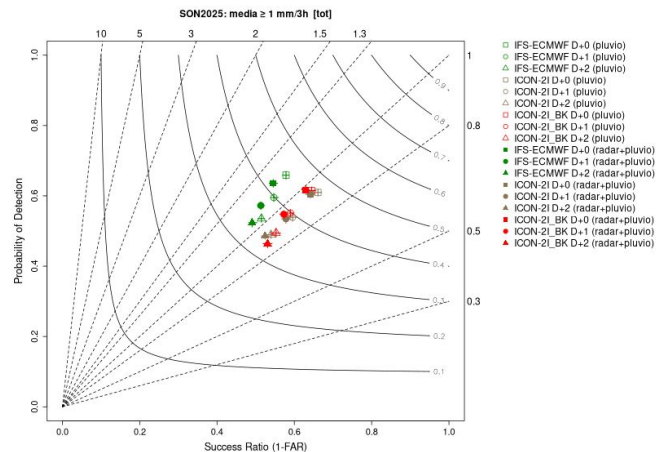
Some Results: mean > 1 mm/3h



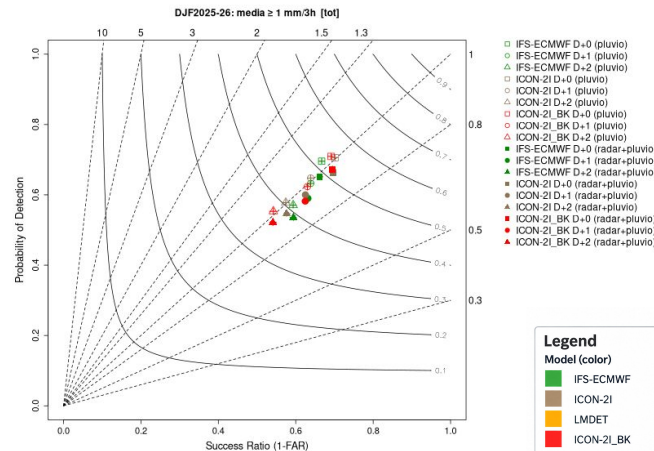
JJA



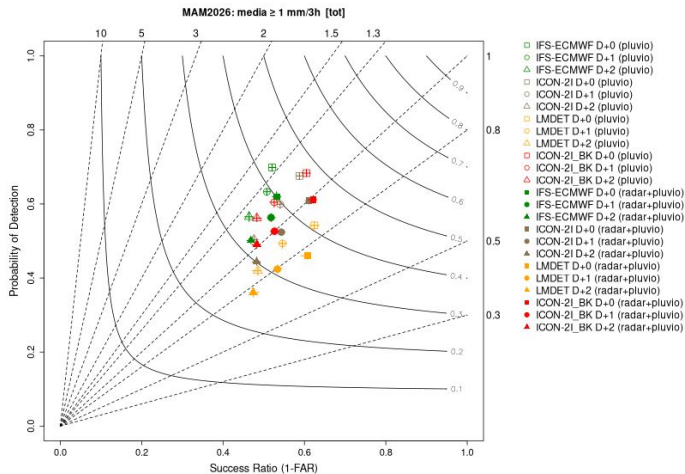
SON



DJF



MAM



Legend

Model (color)

- IFS-ECMWF
- ICON-21
- LMDET
- ICON-21_BK

Lead time (symbol)

- D+0
- D+1
- D+2

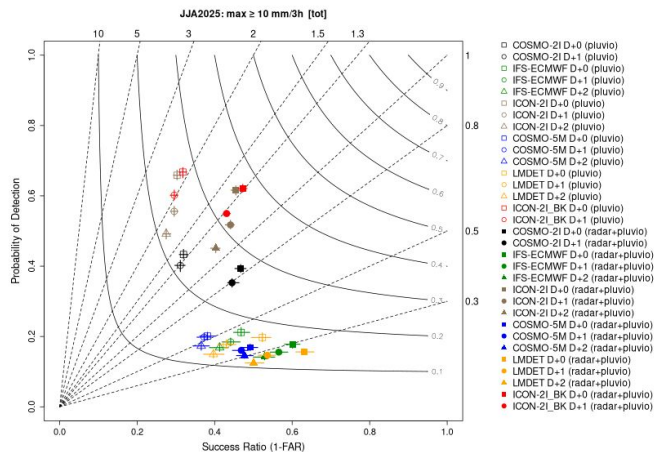
Verification data (fill)

- Rain gauges
- Radar + rain gauges

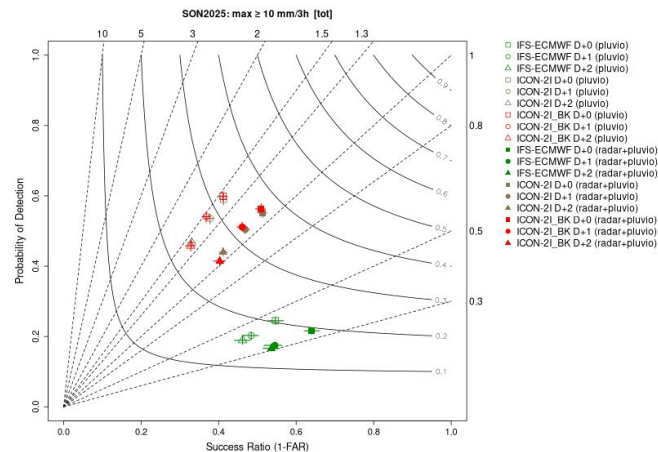
Some Results: max > 10 mm/3h



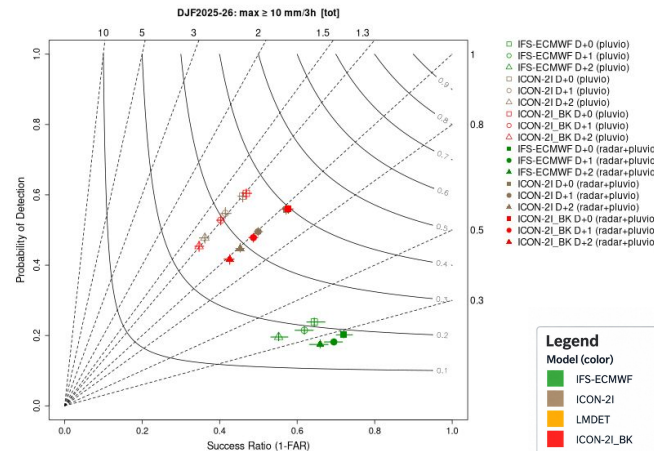
JJA



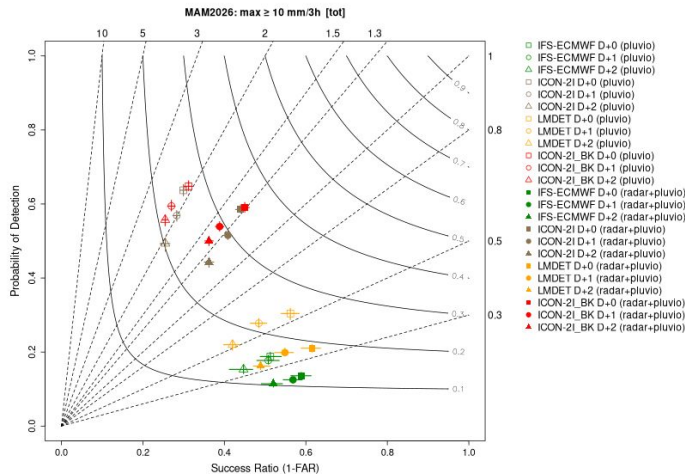
SON



DJF



MAM



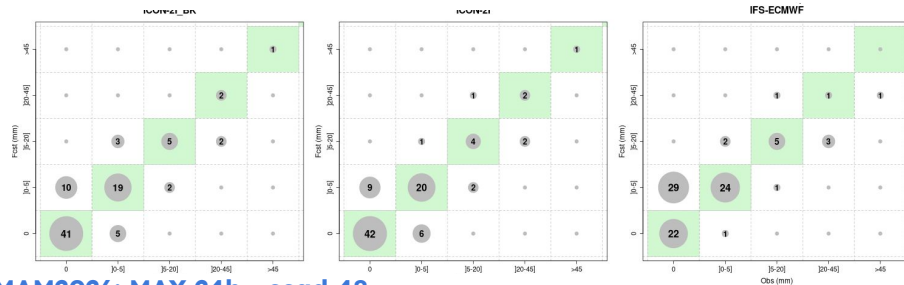
Precipitation - User oriented multi-category verification

Bubble plot: for each warning area, forecast and observed precipitation are assigned to intensity classes and compared visually.

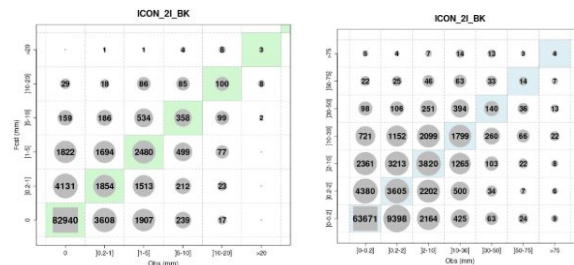
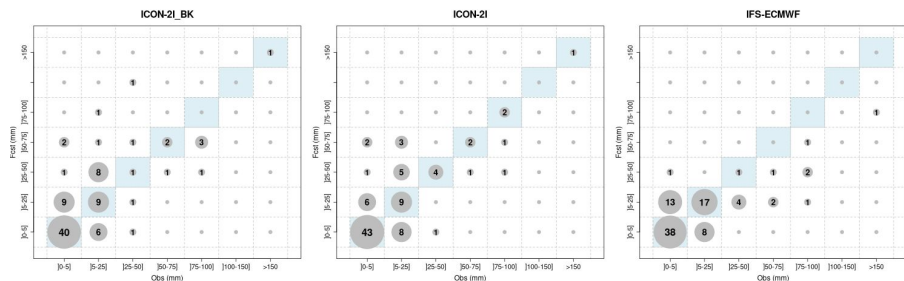
- Diagonal: correct class
- Above diagonal: overestimation
- Below diagonal: underestimation



MAM2026: MEAN 24h - scad 48



MAM2026: MAX 24h - scad 48

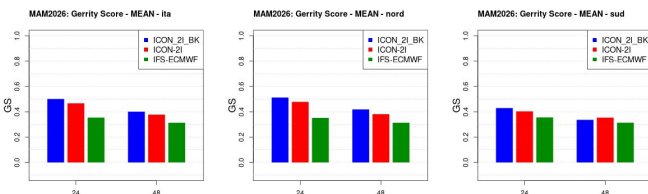


When all warning areas are aggregated, the **Gerrity Score** provides a single summary measure of forecast skill across multiple precipitation classes. It rewards correct forecasts and penalizes errors increasingly as the distance between forecast and observed classes grows.

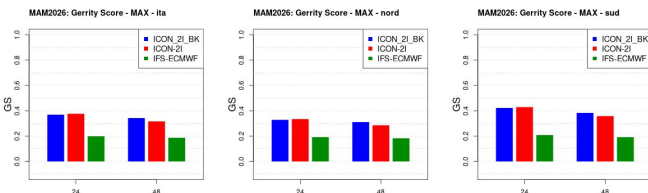
Some Multi-category results

Gerrity Score for 3-h accumulated precipitation, computed using the same mean and maximum intensity classes as in the bubble plots.
Lead times are aggregated over **day 1 (up to +24 h)** and **day 2 (up to +48 h)**.

MAM2026: MEAN



MAM2026: MAX

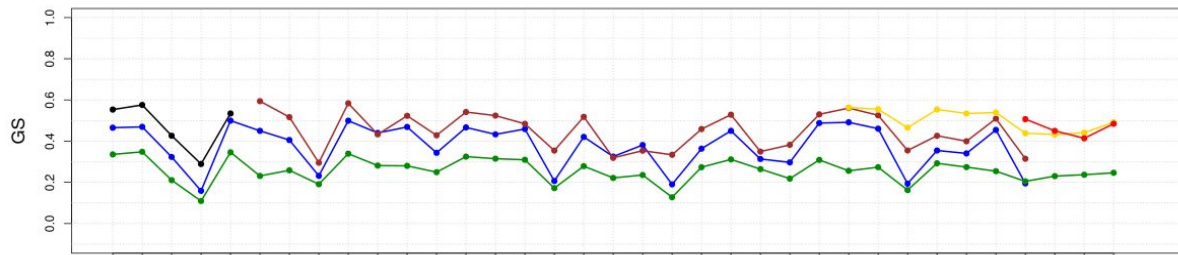


Italy

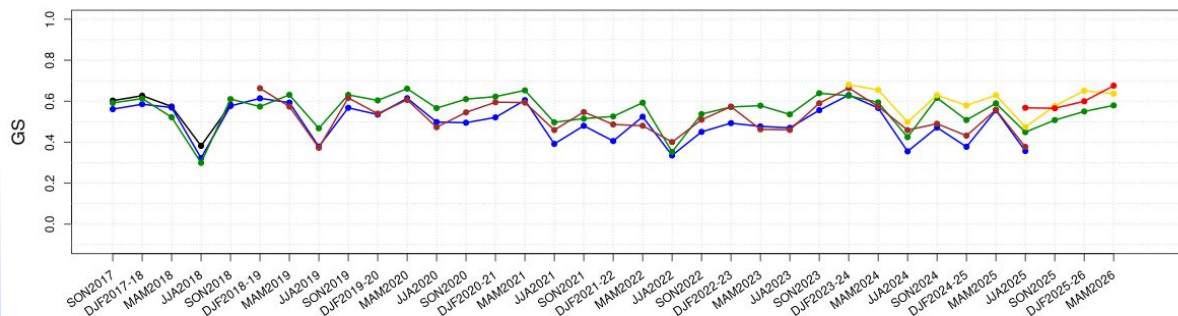
North

South

Gerrity Score for max prec - FC:+48



Gerrity Score for mean prec - FC:+48



—●— COSMO-I2 —●— COSMO-5M —●— IFS-ECMWF —●— COSMO-2I —●— ICON-2I —●— ICON-2I_BK

Seasonal Gerrity Score for 24-h accumulated precipitation at +48 h, using the same mean/max classes as the bubble plots, from SON 2017 to MAM 2026.

Where We Are and What Comes Next

CURRENT STATUS

FFV2 verification

The verification tools are now in place for deterministic and ensemble systems.

CURRENT STATUS

Spatial precipitation verification

FSS and warning-area-based verification provide complementary views of model performance, from spatial skill to operationally relevant precipitation categories.

NEXT STEP

Applying the framework

The focus now shifts from building the verification framework to using it systematically to identify model weaknesses and support model improvement, especially for ensemble and high-impact precipitation forecasts.

THANK YOU!