Highlights of verification activities in COSMO consortium

F.Gofa, A. Iriza Burca, F. Fundel, N. Vela, S. Gabrian, S. Dinicila, M.S.Tesini, F. Batignani, D. Boucouvala, P. Kaufmann, M. Grzelczyk, A. Surowiecki, P. Khain, M. Bogdan,….

WG Verification Activities

- Software: MEC/FFV2, spatial/VAST, EPS: PPCARMENS
- Observations: Crowd source data/PP-APOCS
- Verification Caveats/Open issues
- Common Verification: Restructure Activity
- Data driven Models Evaluation

Felix Fundel,DWD
Deutscher Wetterdienst
Wetter und Klima aus einer Hand

MEC+FFV2

MEC Updated versions ((H. Anlauf))

- **Related to observations:** Humidity, Wind, Global Radiation
- **AI Models:** Support for AI Models on reduced set of model levels (e.g. AICON)

FFV2 updated versions (F.Fundel)

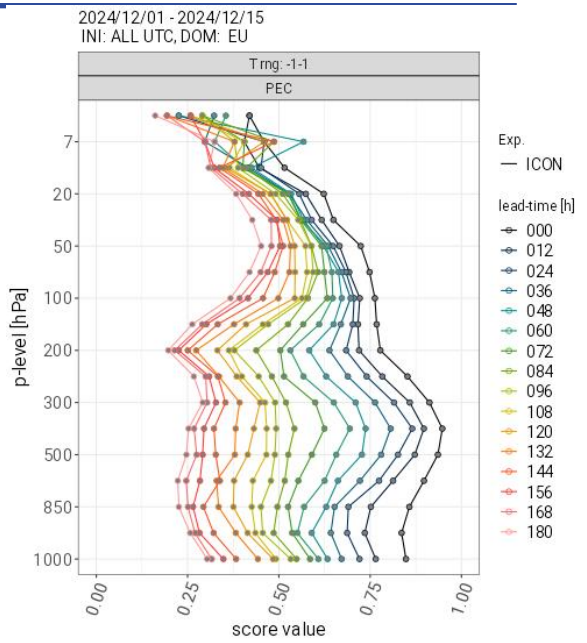
- **Treatment of 10m Winds (INCLUDING** optional filters for wind based on sso_stdh and for too large wind differences between obs and model (>40m/s).
- **Rel. Humidity over Ice RH_i & Categorical Verification of Vertical Profiles**
- **New Scores**
  - **Threshold Weighted CRPS:** Proper ensemble score with weight on (rare) events
  - **Activity:** Quantify forecast and observation activity (spatial variability)

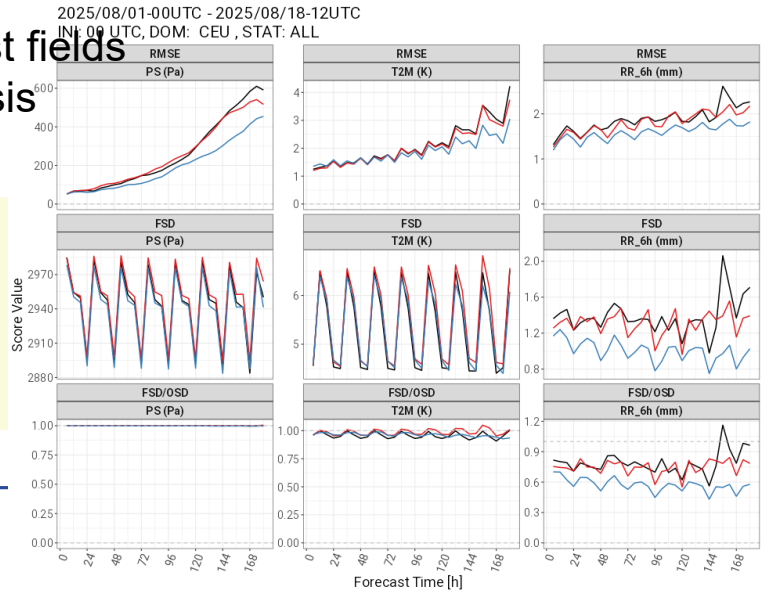
    Smoother fields will be reflected by a reduced std. dev. of the forecasts

    Reduced activity can explain a reduction in RMSE

    Relevant for current AI models that show reduced RMSE due to smooth forecast fields

    Still, a more detailed view on activity is important, e.g. a scale dependent analysis using power-spectra





MEC, FFV2 and Shiny Apps on dkrz DACE gitlab

- Discontinuation of code management on my private gitlab
- https://gitlab.dkrz.de/dace/feedback_file_verification
- All COSMO partners should have access

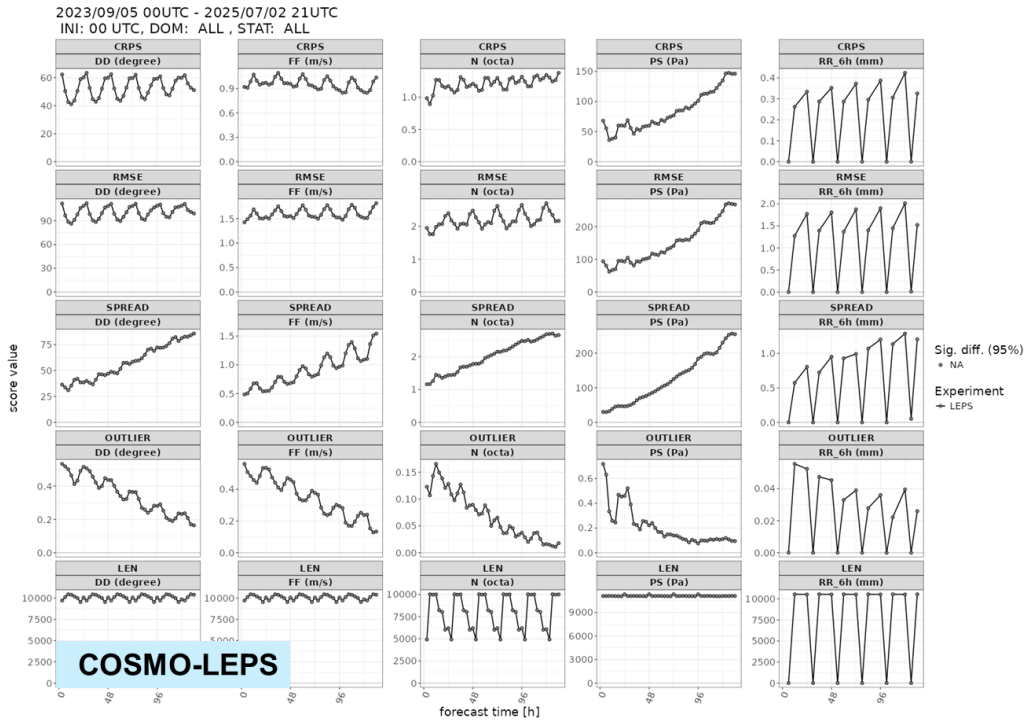
PP CARMENS

Objectives

- Each participating service to set up a web platform that scores based on local EPS system outputs, are calculated (MEC+FFV2) and visualized.
- All the applications of the system, mec, ffv2, shiny currently have the ability required to fulfil the project.
- An EPS dedicated shiny platform on COSMO web that the stats (rdata) files from each EPS system plus COSMO-LEPS will be uploaded at least for a season following all the prerequisites of CP verification.
- After the completion of the PP this infrastructure will be used for Common Verification activity.

PP-Leader Ioan-Stefan GABRIAN

with contributions from F. Fundel, F. Gofa, D. Boucouvala, T. Andreadis (HNMS), F. Batignani (CoMET), P. Khain, (IMS), A. Pauling, Pirmin Kaufmann (MCH), S. Dinicila (NMA) , Enrico Minguzzi(Arpae), Thomas



Management and problem solving
https://gitlab.dkrz.de/cosmo/pp-carmens

III. Verification Caveats

WMO-No. 8

Guide to Instruments and Methods of Observation (WMO-No. 8)

5.8.2 Peak gusts and standard deviation

Before specifying the appropriate response characteristics of wind-measuring systems, it is necessary to define the gust duration as required by the application. Wind extremes are mainly used for warning purposes and for the climatology of extreme loads on buildings, constructions and aircraft. It is important to realize that the shortest gusts have neither the time nor the horizontal extent to exert their full damaging effect on large constructions. WMO (1987) concludes that a gust duration of about 3 s accommodates most potential users. Gusts that persist for about 3 s correspond to a "wind run" (duration multiplied by the average wind speed) of the order of 50 to 100 m in strong wind conditions. This is sufficient to engulf structures of ordinary suburban/urban size and to expose them to the full load of a potentially damaging gust.

——, 1987: *The Measurement of Gustiness at Routine Wind Stations: A Review* (A.C.M. Beljaars). Instruments and Observing Methods Report No. 31. Geneva.

However…

Country	NWS	Gust Definition
Germany	DWD	3 s (to be confirmed)
France	MeteoFrance	3 s (to be confirmed)
Austria	GeoSphere	2 s
Italy	Meteo Alto Adige	3 s (except mountain peaks: 1 s)
Italy	ARPA Lombardia, ARPA Piemonte	5 s
Italy	Centro Funzionale Valle d'Aosta	1 s
Switzerland	MeteoSwiss	1 s

Pirmin.Kaufmann@MeteoSwiss.ch

Measured Frequencies of Occurrence

All Swiss Stations; Winter (DJF) 2023/24, Summer (JJA) 2024, Winter 2024/25

Score	Season	3s Gusts (WMO)	1s Gusts (MCH)	Relative Change 3s → 1s
Frequency of **medium** gusts ≥ 12.5 m/s (≈ 25 kn)	Wi 23/24	9.9%	11.0%	11% increase
	Su 2024	2.90%	3.50%	21% increase
	Wi 24/25	7.6%	8.4%	11% increase
Frequency of **strong** gusts ≥ 20 m/s (≈ 70 km/h)	Wi 23/24	1.88%	2.27%	21% increase
	Su 2024	0.227%	0.295%	30% increase
	Wi 24/25	1.36%	1.65%	21% increase
Frequency of **very strong** gusts ≥ 25 m/s (= 90 km/h)	Wi 23/24	0.612%	0.769%	26% increase
	Su 2024	0.058%	0.072%	24% increase
	Wi 24/25	0.367%	0.493%	34% increase

Pirmin.Kaufmann@MeteoSwiss.ch

Freq. Medium Gusts

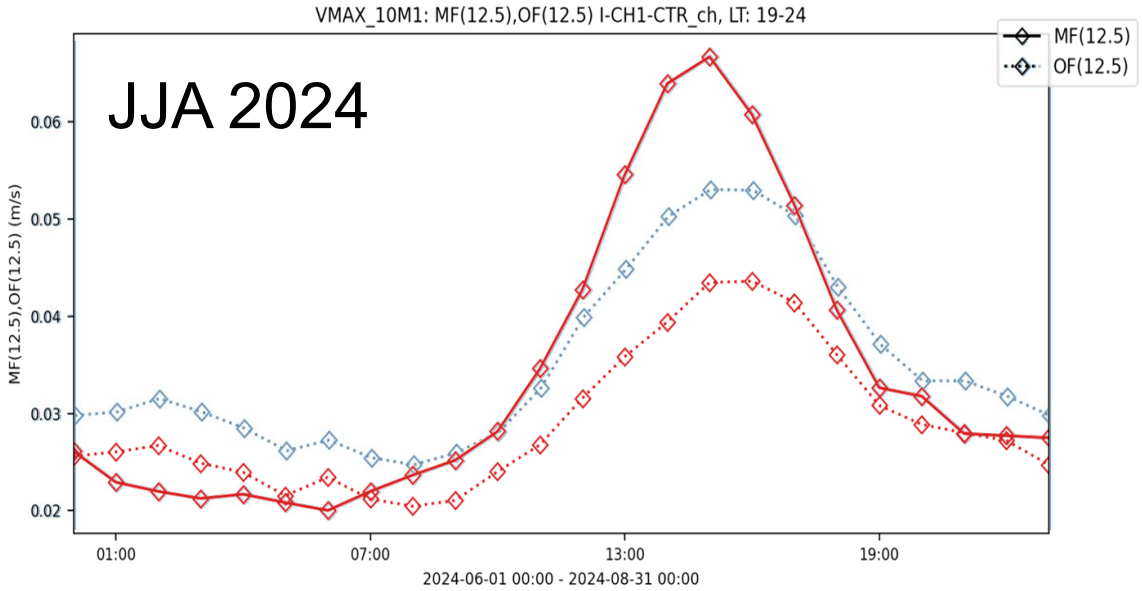
Diurnal Cycle

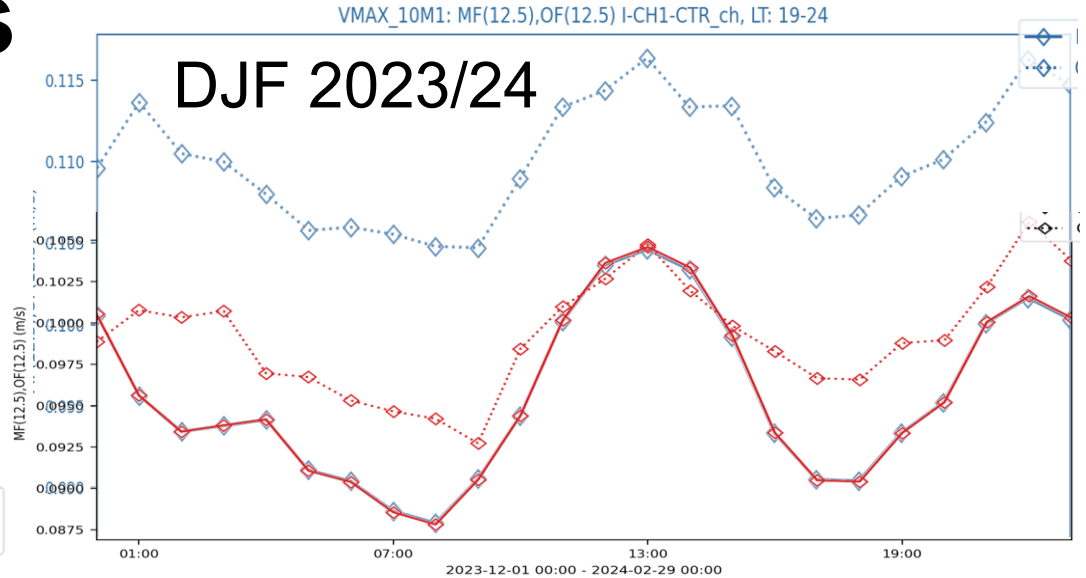
Gusts > 12.5 m/s (≈ 25 kn)

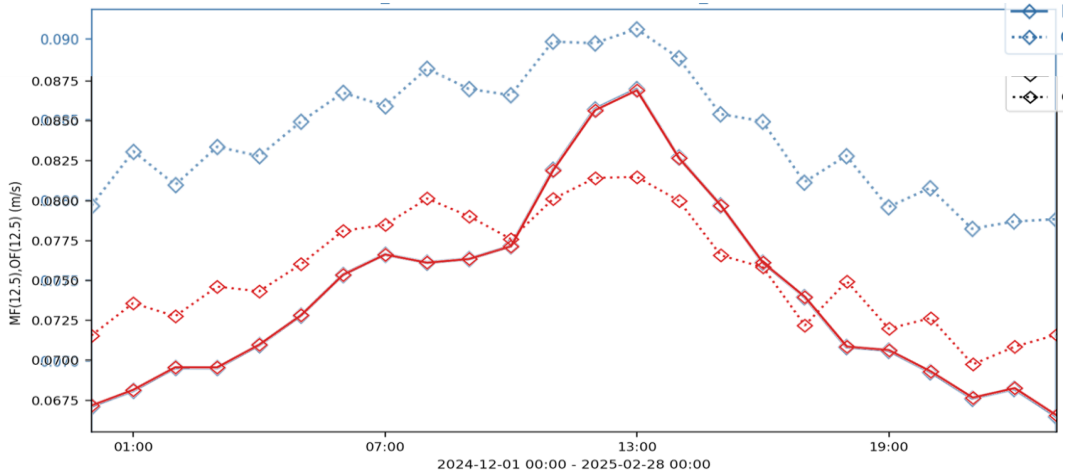
◇—◇ ICON-CH1-EPS

◇···◇ Obs: 1s gusts

◇···◇ Obs: 3s gusts







DJF 2024/25

Pirmin.Kaufmann@MeteoSwiss.ch

Change in ICON-CH1-EPS Scores

Score	Sea.	Change 1s → 3s Gusts	Relative Change
Frequency Bias ≥12.5 m/s	DJF JJA DJF	0.87 → 0.96 0.97 → 1.17 0.88 → 0.98	Decrease of underestimation (9% less) Overestimation appears (17%) Underestimation vanishes (10% less)
Frequency Bias ≥20 m/s	DJF JJA DJF	0.81 → 0.98 0.89 → 1.16 0.89 → 1.08	Decrease of underestimation (17% less) Change from under- to overestimation Change from under- to overestimation
Frequency Bias ≥25 m/s	DJF JJA DJF	0.90 → 1.13 1.06 → 1.34 1.03 → 1.38	Change from under- to overestimation Increase of overestimation (28% more) Increase of overestimation (35% more)
Eq. Threat Score ≥12.5 m/s	DJF JJA DJF	0.44 → 0.44 0.25 → 0.24 0.44 → 0.44	No change (similar for other thresholds and for Threat Score)

Pirmin.Kaufmann@MeteoSwiss.ch

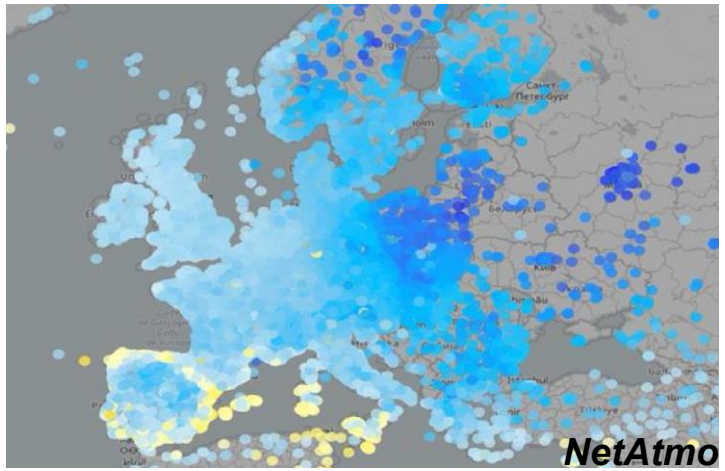
Priority Project: APOCS (Application of Personal Weather Station and Opportunistic Sensor Data CrowdSourcing)

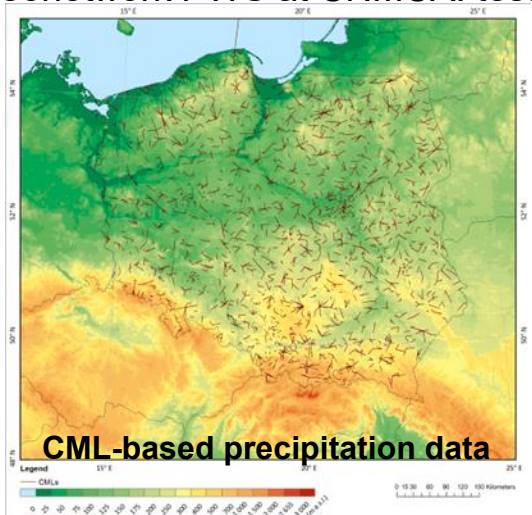
IMGW-PIB: Artur Surowiecki, Andrzej Wyszogrodzki, Witold Interewicz, Grzegorz Zakrzewski, Jan Szturc, Anna Jurczyk, Katarzyna Ośródka, Magdalena Szaton, Radosław Drożdżoł, Bartłomiej Sobczyk, Adam Jaczewski
HNMS: Flora Gofa, Dimitra Boucouvala
CIMA: Massimo Milelli, Elena Oberto, Francesco Uboldi
ARPA Piemonte: Valeria Garbero
Politecnico di Torino: Tanguy Houget
CNMCA: Valerio Cardinali

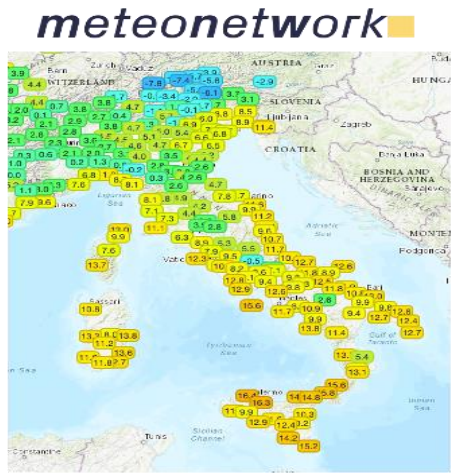
II. Observations: Crowd source data

AIM: application of alternative weather data acquired by the Personal Weather Stations (PWS) and other Opportunistic Sensors (OS) into research and operations activities at the level of national weather services.
FOCUS: building up useful database with PWS for research purposes; application and testing of the developed RainGaugeQC and Titan-Quality Control (QC) software packages and evaluation of the quality of different physical parameters including precipitation, temperature and humidity; and the application of the products developed from the PWS data to model assimilation and numerical forecast verification.

Tasks: Development of COSMO PWS database, building up parser for decoding data formats, adaptation or building up transfer protocols, development database functionalities. Extended testing for different QC algorithms: Adaptation of the RainGaugeQC system to CML-based precipitation, Titan lib applied to "not conventional" raingauge datasets, development of QC procedure for NetAtmo observational data. Application of PWS-based gridded products: spatial verification techniques, application to NWP model verification in urban areas, use of gridded RainGRS+ precipitation estimates for assimilation (COSMO-RUC), Meteonetwork PWS at CNMCA: testing stability and consistency.







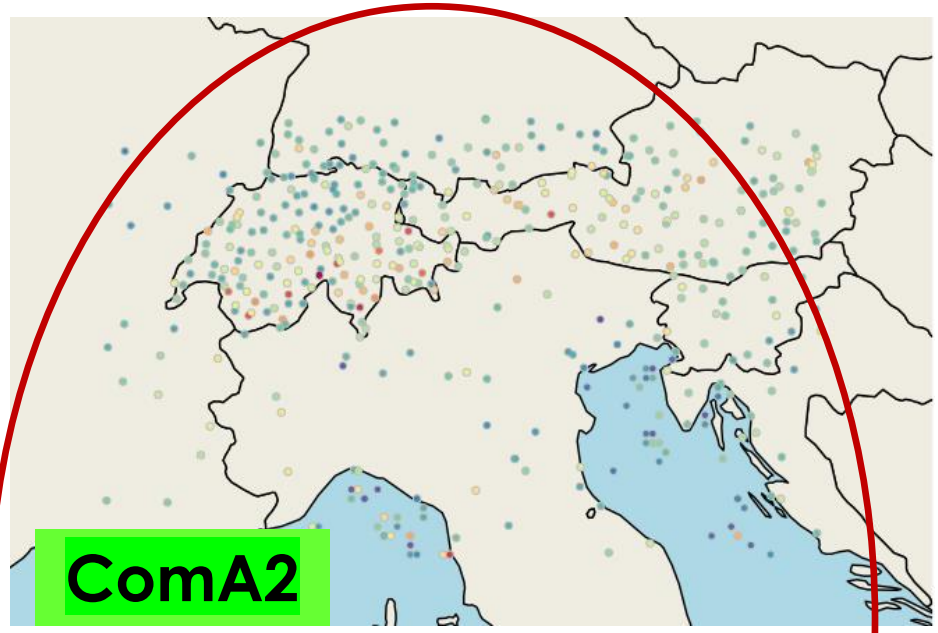
Common Plots Activity Restructure

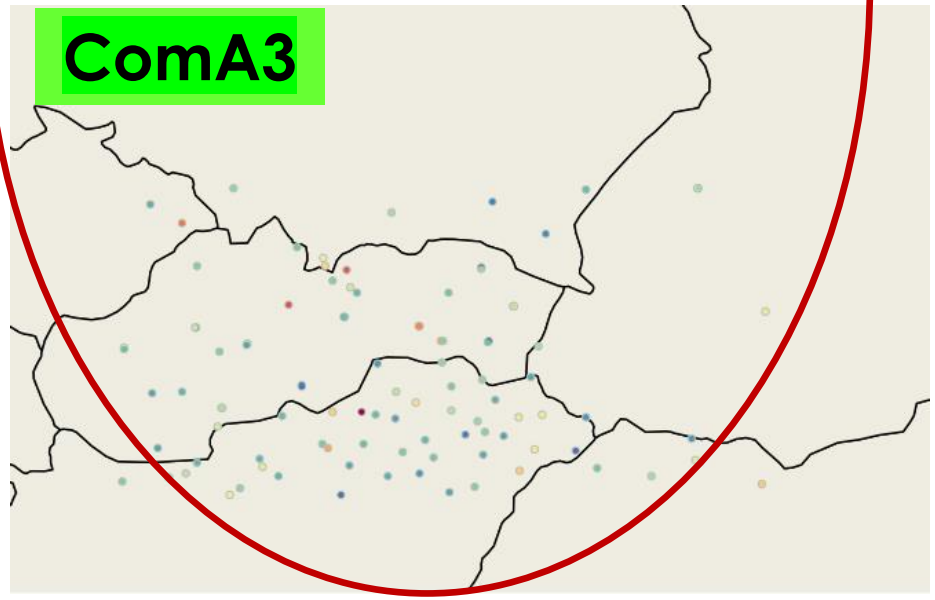
The activity concerns the calculation and representation of verification results using the operational ICON-LAM implementations in each service, both det and eps with MEC-FFV2 platform

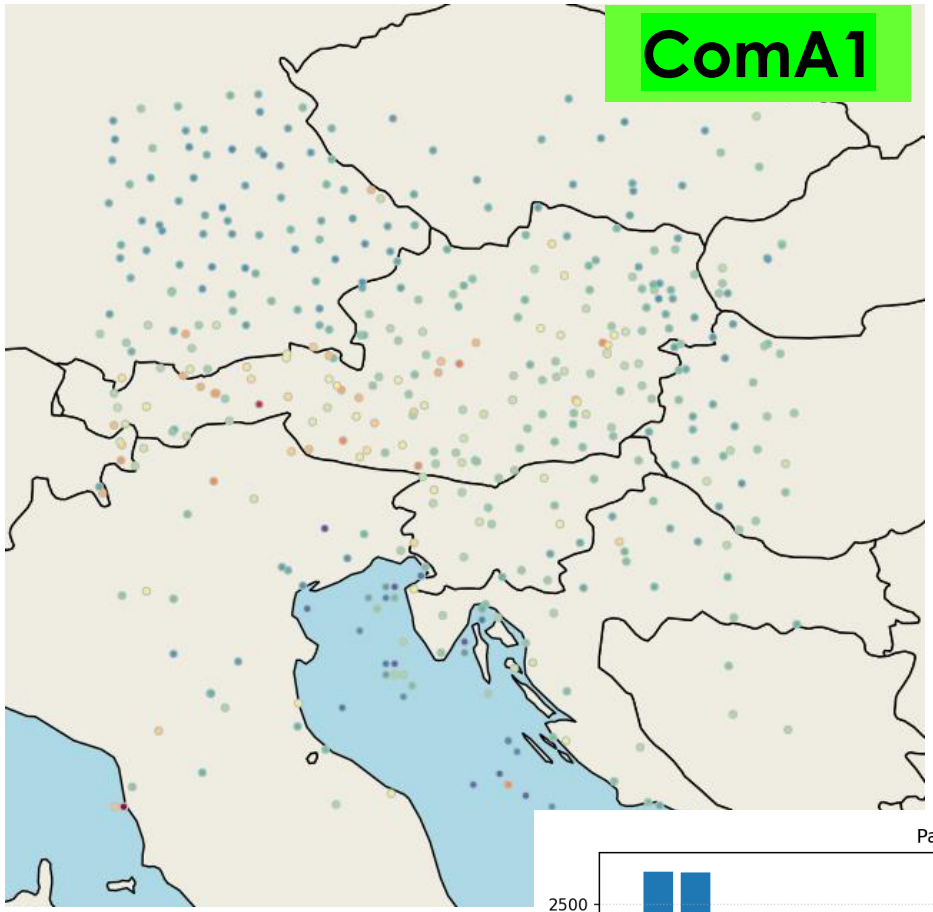
Reasoning:

- Common Plot verification activity has steadily expanded, both in terms of scope and complexity.
- Integration of deterministic (limited area and global), ensemble models and additional statistical requirements has increased the operational workload
- Important to reformulate and reassign roles to ensure the sustainable continuation and quality of the CP activity.

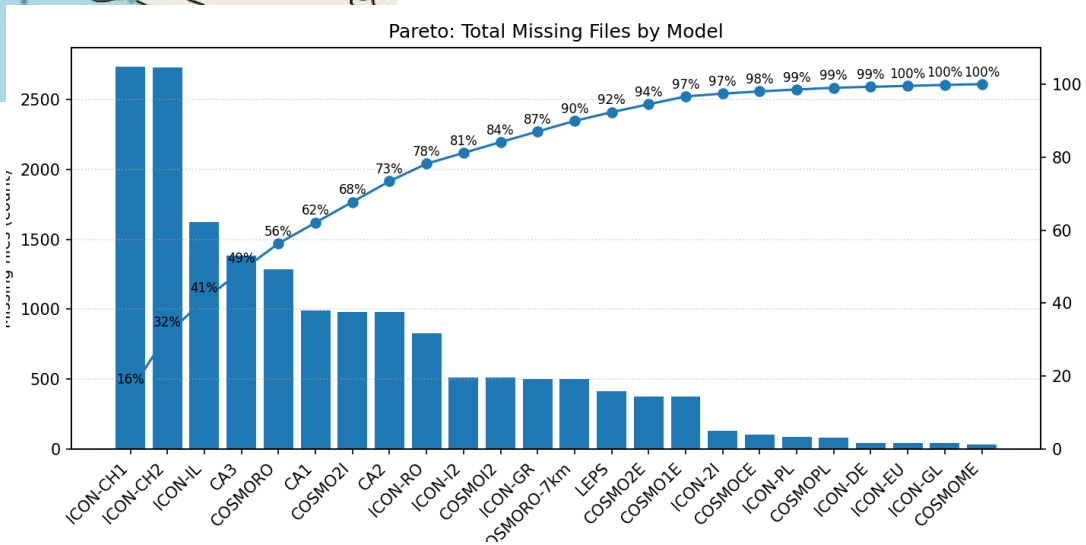
V. Common Verification:Restructure, Model Errors







S. Gabrian, NMA
Common Plots

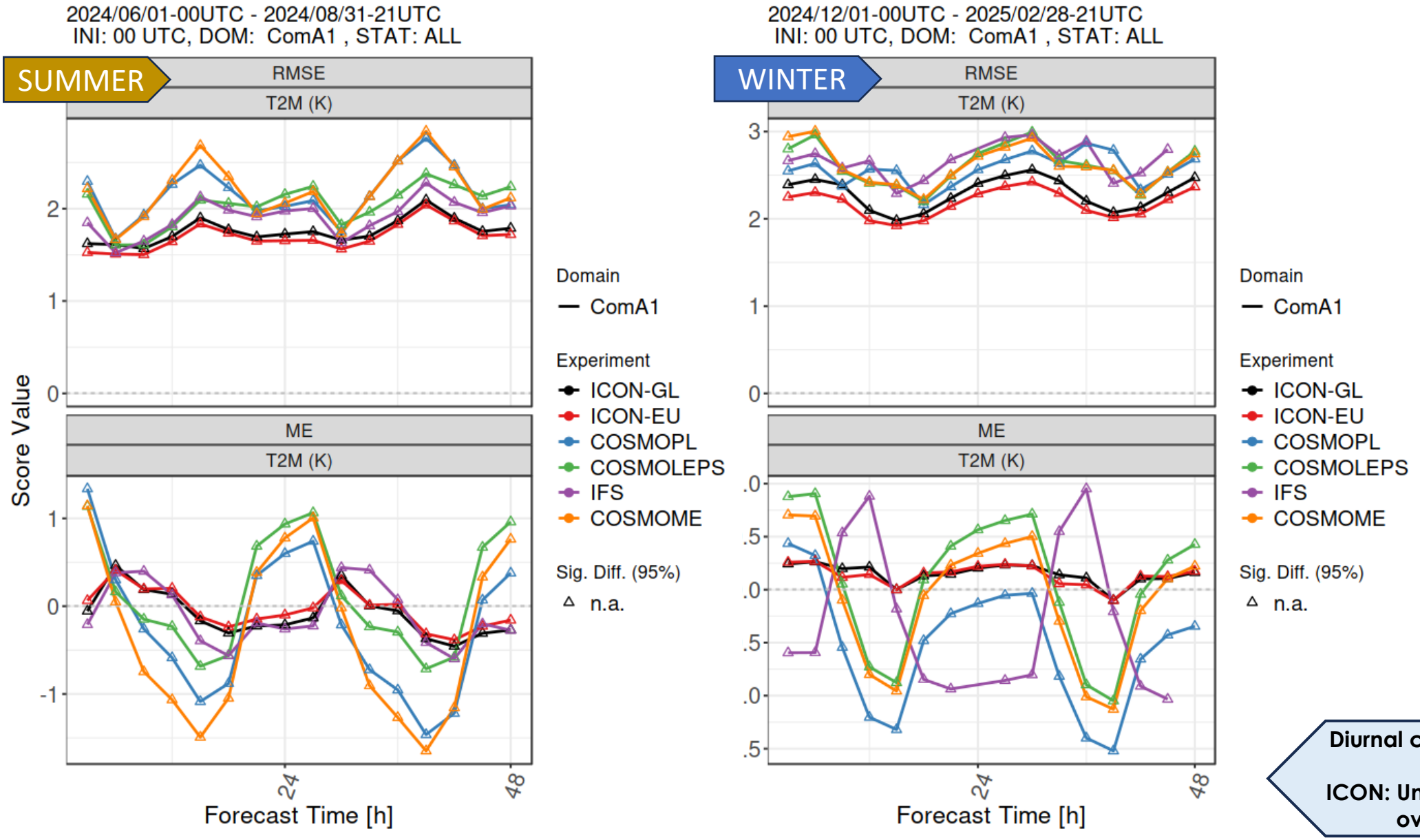


CONSORTIUM FOR SMALL SCALE MODELING
COSMO

Driving Models

ComA1

2mT

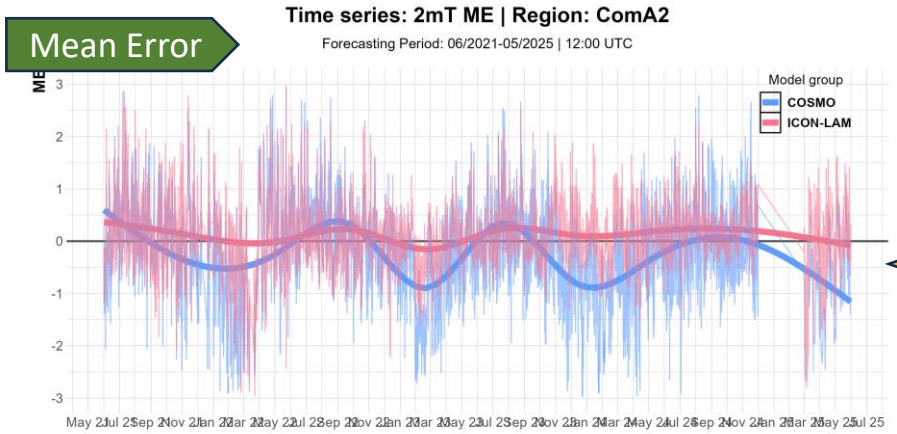


Diurnal cycle stronger in COSMO+IFS
Opposite phase
ICON: Underestimation @night mainly, overestimation @midday

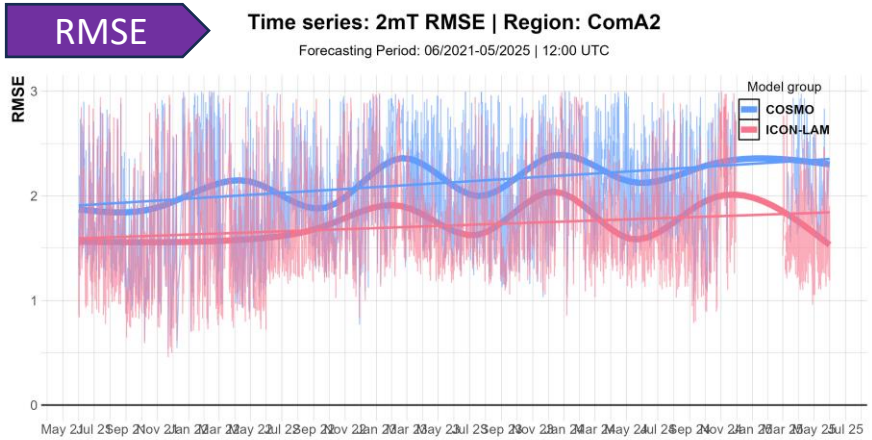
2mT @12+24UTC: HRES COSMO/ICON ComA2. 2021-2025

Factorial timeseries linked to ME and RMSE, thicker lines represent the smoothed average. Used: geom_smooth() that adds a regression line to a plot, and it uses a loess smooth when there are fewer than 1000 observations, and a GAM when there are more.

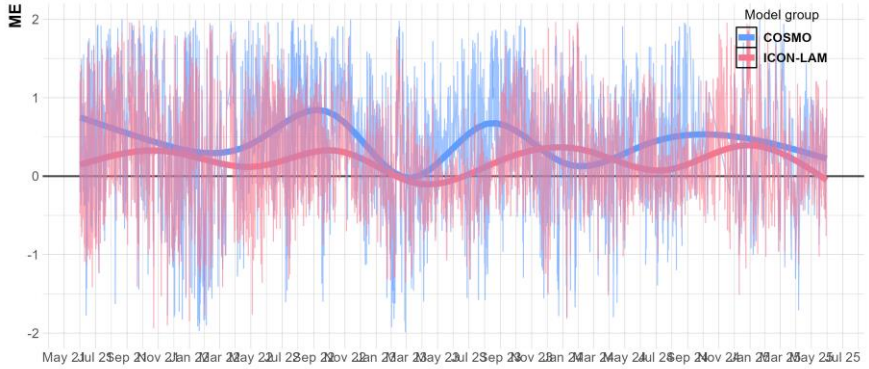
Mean Error



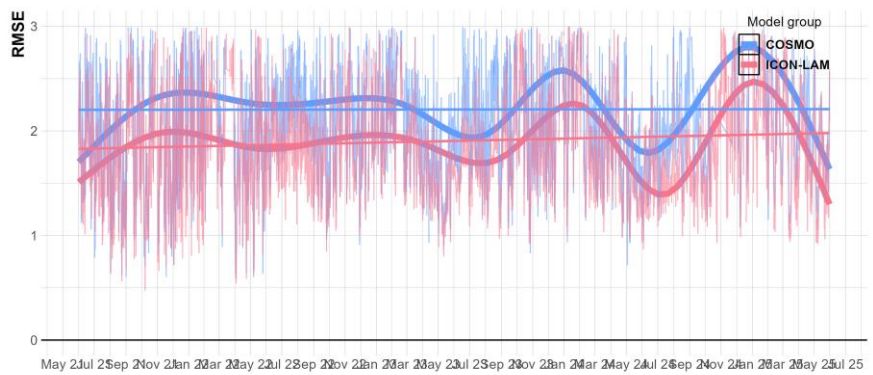
RMSE



Mean Error



RMSE



2 m T

SYS: RMSE diurnal cycle present in all models. Large errors in run initialization. Reduced RMSE/Bias DC with ICON

SYS : Larger errors at night and early morning in winter and at midday in the summer.

SYS : Underestimation during summer warm hours. Overestimation during summer at night

SYS : ICON tendency to underestimate mainly in the winter night hours

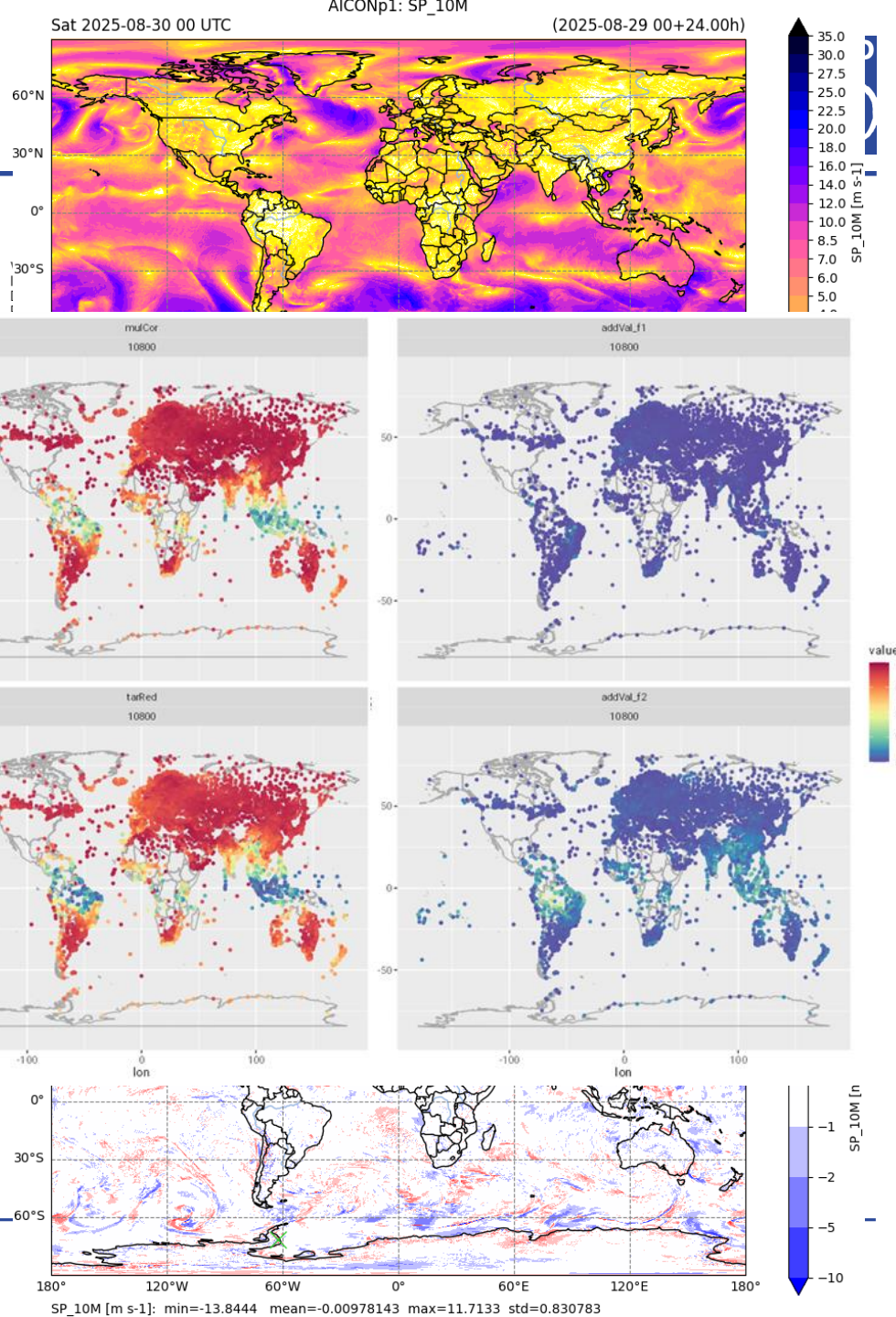
Possible Model Error attributes: issues in the surface energy budget, representation of land-atmosphere interactions, errors in cloud cover or radiation schemes, incorrect representation of soil moisture or boundary Layer parameterization.

Newsletter section on Models Performance
https://www.cosmo-model.org/content/model/documentation/newsLetters/default.htm

VI. Data driven Models Evaluation

AICON

- AICON nternally operational this week
  - More intense evaluation also by forecasters
  - Limited output FF10M, T2M, TD2M, RH2M, TOT_PREC and U, V, RH, T, Z on 13 model levels (close to standard pressure levels)
  - Some artefacts like unphysical values or unrealistic precipitatio at some grid cells need to be addressed
- AICON verification
  - Standard scores look promising
  - Additional methods and metrics are under development
  - Scale dependent activity (spatial variance) (Britta Seegebrecht)
  - Simple, observation based activity estimate in FFV2 (Felix Fundel)
  - Additionally explained variance after Glowienka-Hense et al. (2020) (Sabrina Wahl)
- Upcoming developments
  - AICON-LAM



Verification of data driven Forecast Models
Challenges & Alignment with NWP Practices

- **Reference Datasets Matter**

NWP: usually against *observations* (radiosondes, SYNOP, satellite)
AI/ML: often against *reanalyses (area)*
Focus: establish *common reference frameworks*

- **Metrics**

RMSE, ACC, CRPS often look "better" for AI (smoothed fields)
Focus: introduce measures to quantify forecast and observation activity, spatial variability (variance, power spectra)

- **Extremes and Physical realism**

NWP can capture extremes but scores worse on average.
Focus: Test conservation (mass, moisture),verify tails with threshold scores, metrics sensitive to extremes, spatial structures & rare events.
Evaluate across regimes, seasons. rare events

Verification of data driven models is not only about performance scores, it is the scientific check that ensures reliability, transparency, and trust in model forecasts

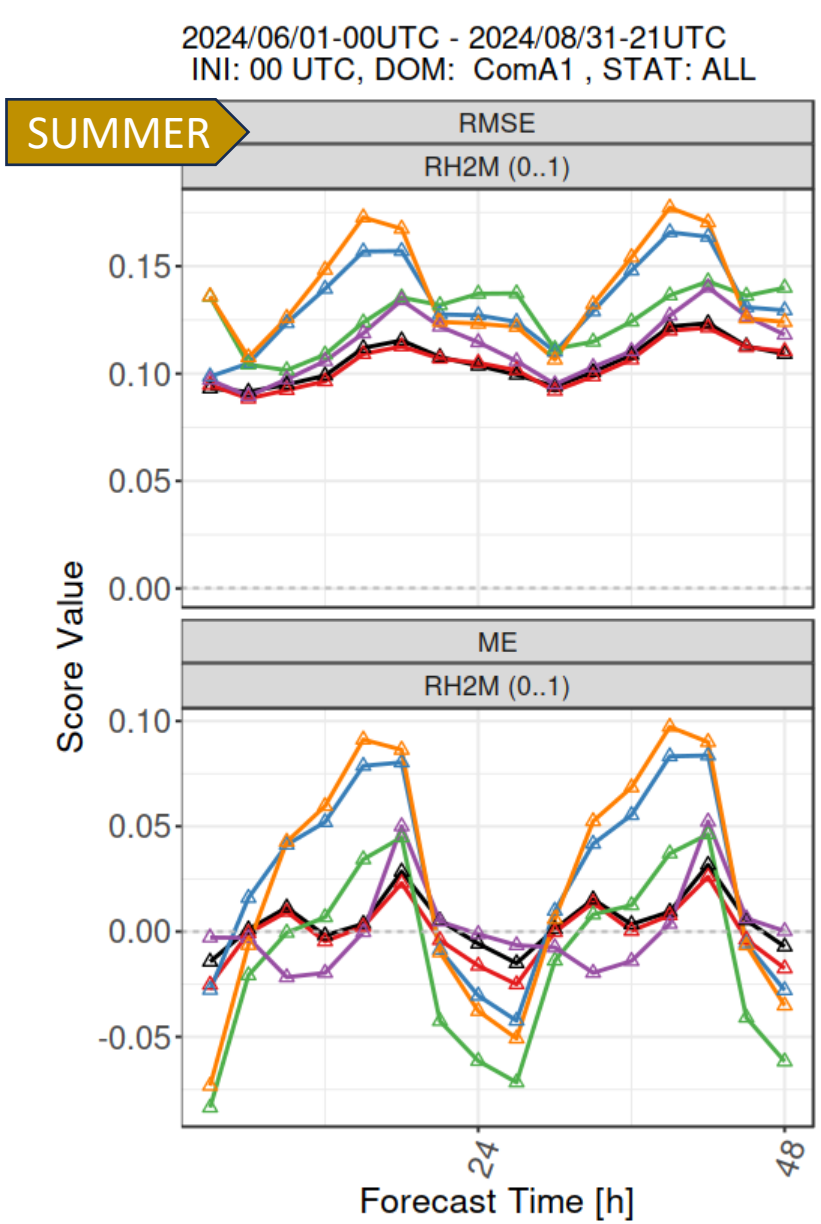
New COSMO Work Group Verification and Applications Co-Chair

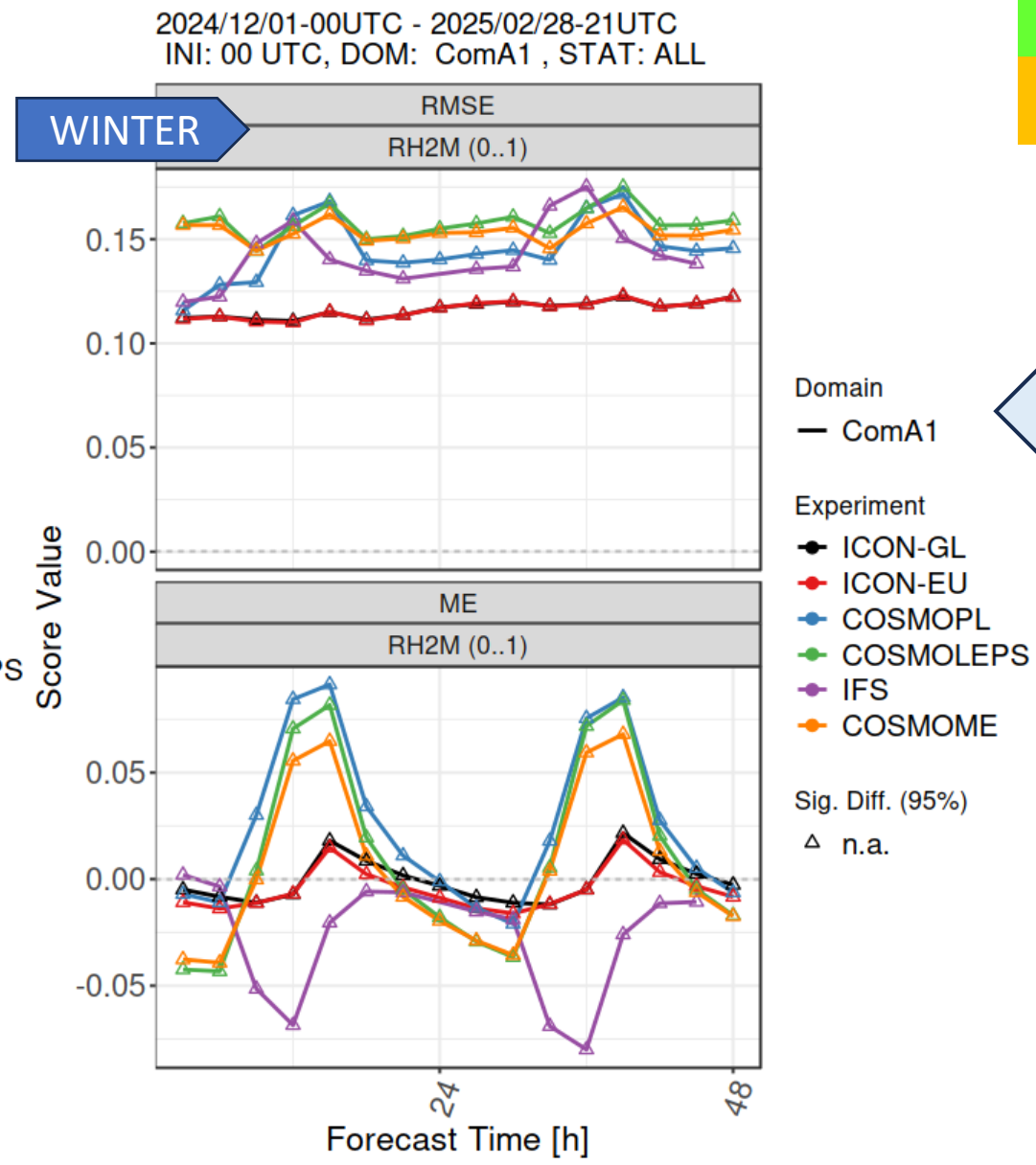
Amalia Iriza-Burca, NMA (starting 01.2026)

Driving Models

ComA1

RH

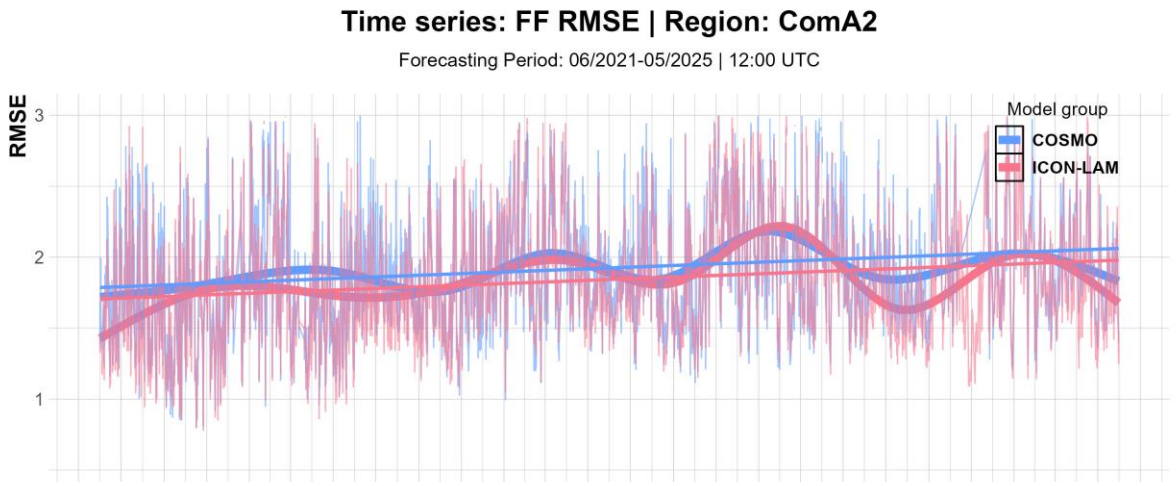




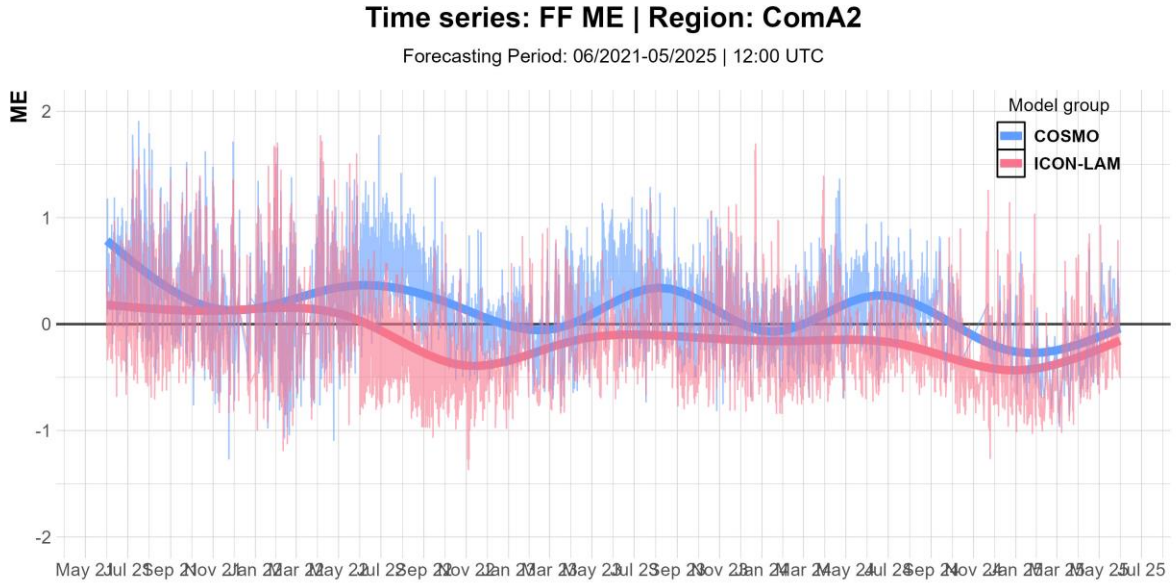
ICON+ IFS in general drier

WinsSp @12UTC: HRES COSMO/ICON ComA2, 2021-2025

Mean Error



RMSE



WindSp

HighRes LAM: No differences in RMSE variability among models. Underestimation mainly from ICON-LAMs

SYS: Larger error cycle in summer

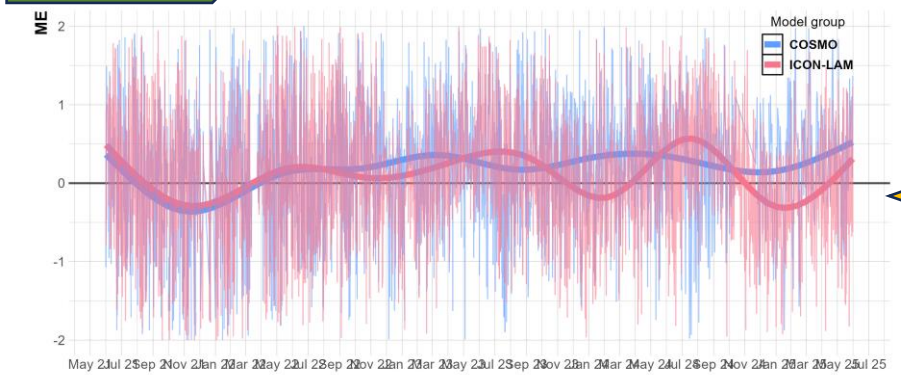
ModDep: negative bias in all seasons with max during nighttime mainly in winter

Possible Model Error attributes: errors in the momentum fluxes or frictional effects in the boundary layer, inaccuracies in the pressure field, particularly in the simulation of high and low-pressure systems (wind direction)

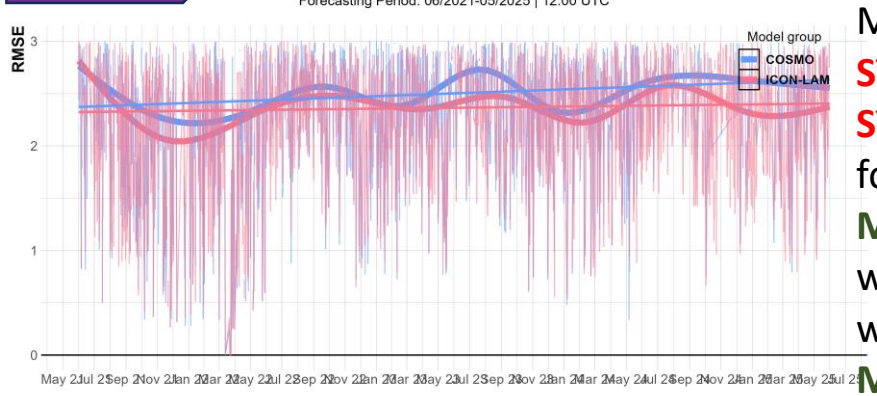
TCC @12+24UTC: HRES COSMO/ICON ComA2, 2021-2025

Factorial timeseries linked to ME and RMSE, thicker lines represent the smoothed average. Used: geom_smooth() that adds a regression line to a plot, and it uses a loess smooth when there are fewer than 1000 observations, and a GAM when there are more.

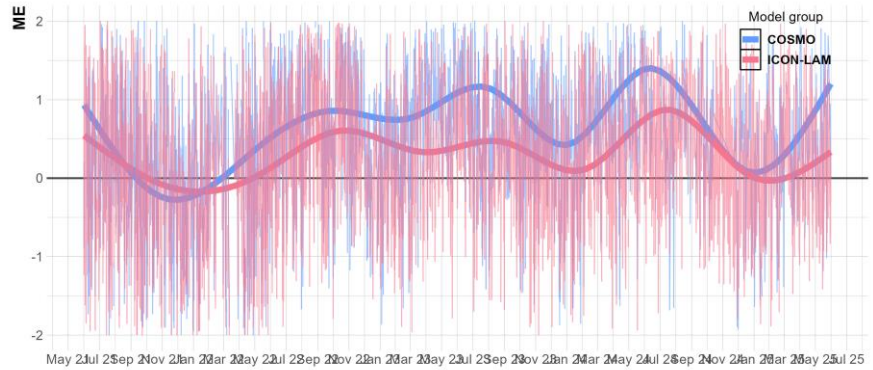
Mean Error



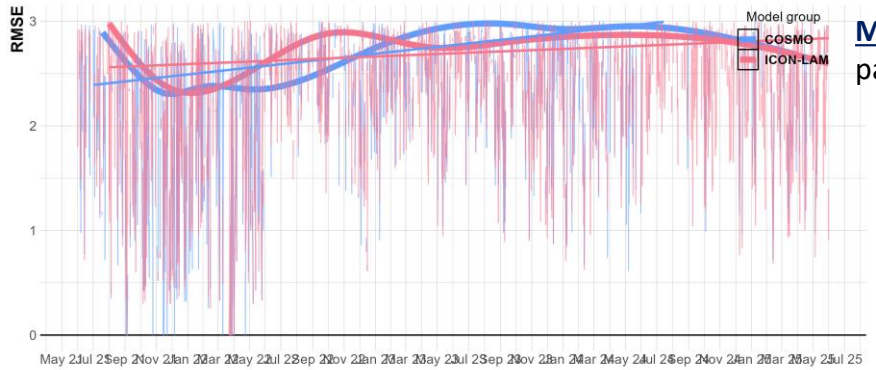
RMSE



Mean Error



RMSE



TCC

SYS : Diurnal cycle of both ME/RMSE for TCC

SYS : **La**rger errors in summer

SYS : Higher errors during nighttime for all models (overestimation)

ModDp: Higher underestimation in warm hours of the day with ICON in winter

ModDp: Ambiguous performance in RMSE (higher errors at night) with a tendency to overestimation during warm months night hours.

Model Error attributes: shortcomings in cloud parameterization, radiation errors

Newsletter section on Models Performance

https://www.cosmo-model.org/content/model/documentation/newsLetters/default.htm