WG Verification and Applications

COSMO Update

F.Gofa, F. Fundel, J. Linkowska, N. Vela, S. Gabrian,
S. Dinicila, M.S.Tesini, F. Sudati, F. Batignani, D. Boucouvala, P. Kaufmann,
M. Zosicz, M. Grzelczyk, P. Khain, B. Maco,…

flora.gkofa@hnms.gr

SOFTWARE

Methods - HIW

Projects - Collaborations

COMMON Verification

Presentation Flow

VAST 3.0

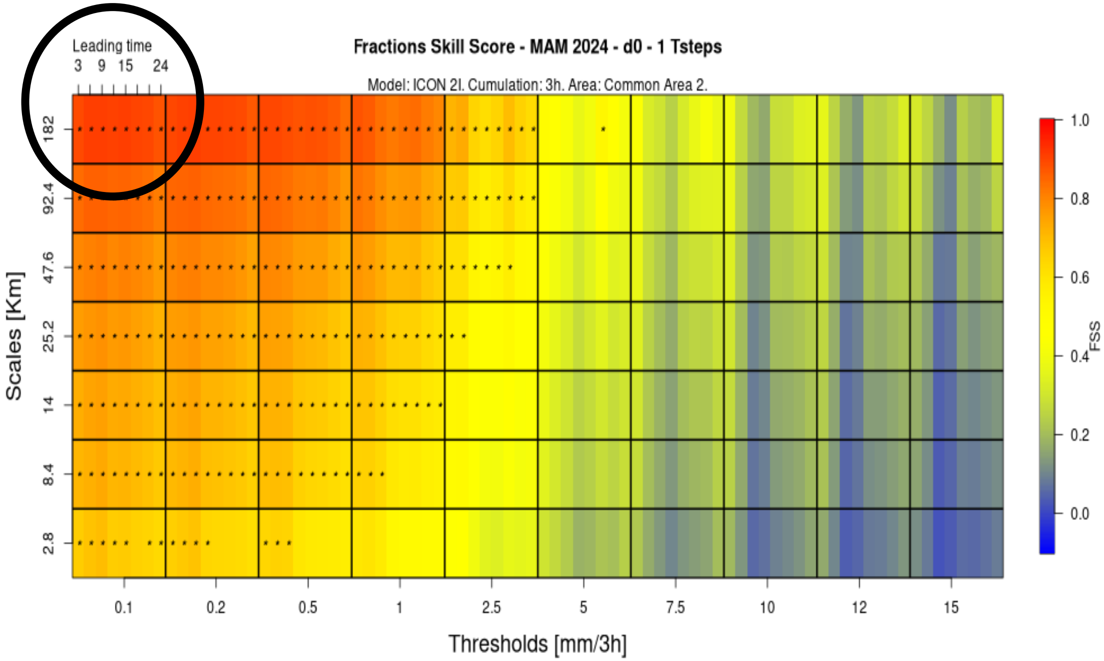
Naima Vela, Arpa Piemonte

VAST *Versus Additional Statistical Techniques* is a software package composed to verify the dichotomic parameters of the ICON-LAM according to a **multi intensity** and **multi scale** spatial verification approach.

Categorical scores can be calculated through **different upscaling approaches** (normal upscale, yes/no, minimum coverage, fuzzy logic, etc.).

Non categorical scores are also included (FSS, BSS and ETSr).

INPUT: GRIB files (forecast and observation)
OUTPUT: heatmap plots according to the user's selection



Technical info

- The grib files are elaborated using **CDO** functions (https://code.mpimet.mpg.de/projects/cdo/wiki/Cdo#Documentation)
- The ECMWF package "**grib_api**" or "**ecCodes**" is needed by the user to elaborate the grib files https://www.ecmwf.int/en/elibrary/80529-grib-api , https://confluence.ecmwf.int/display/ECC
- Part of the fuzzy elaboration is run through Fortran 90 scripts compiled using the **gfortran** compiler (https://gcc.gnu.org/onlinedocs/gfortran/)
- The plots are created by VAST using **R** (https://www.r-project.org/) A version **3.0** or higher is required.

Advances in Feedback File based Verification

Felix Fundel
Deutscher Wetterdienst
Wetter und Klima aus einer Hand

FFV2 (statistical module) Update

- Allow upper and lower threshold in ensemble verification
  - e.g. categorical and probabilistic scores for gusts between 10 and 20 m/s
- Same Tmin/Tmax observation filter consistent with TD2M and T2M
  - z_station – z_modsurf <=150m
  - sso_stdh <= 200
- Updating bitcheck function (bitcheck2)
  - bitcheck allows to filter bit encoded variables (e.g. ‚flags‘ that hold reasons for obs. rejection)
- Additional filtering options
  - ‚mdlsfc‘ (model surface characteristics)
  - ‚level_sig‘ (type of TEMP level)
- Extension for ocean observations (ARGOS, OSTIA, SMOS)

MEC (FF preparation) Update

AI model verification needed

- DWD started developing an own AI model (AICON) and reanalysis (DREAM) for training
- Available AI models (Pangu, GraphCast, AIFS.. run by ECMWF) have reduced output
- T, Z, U, V, RH on 13 pressure levels and T2M, U10M, V10M, TD2M, PMSL, (TOT_PREC) on 0.25° ERA5 grid

MEC adaptations

- Needed to cope with fewer levels (linear vertical interpolation to avoid overshooting)
- Needed to works with PMSL and PS
- But still uses lowest model level for FF10M, renders verification on pressure levels useless
- MEC developments by Harald Anlauf

Still some MEC work coming up with AICON

MEC needs to work on model levels without HHL. For FF10M verification MEC needs to use surface wind not lowest model level

Advances in Feedback File based Verification

FFV2 (statistical module)

Revised Documentation
http://www.cosmo-model.org/shiny/users/fdbk/RfdbkVeriDoku.html

Some updates relevant for EPS verification

Shiny Apps (visualization platform)

- Apps require a config file FFVSHINYAPPCONFIG
  - Sets some site specific variables needed in all apps
  - E.g. sever name, path to score files, app directory
  - Allows to port apps more easily to new environments
- Namelist in extra tab
  - The verification namelist is shown in the app
  - Helps to follow verification setup and debugging
- New gitlab repository
  - https://gitlab.com/rfxf/ShinyApps
  - Including a short setup instruction

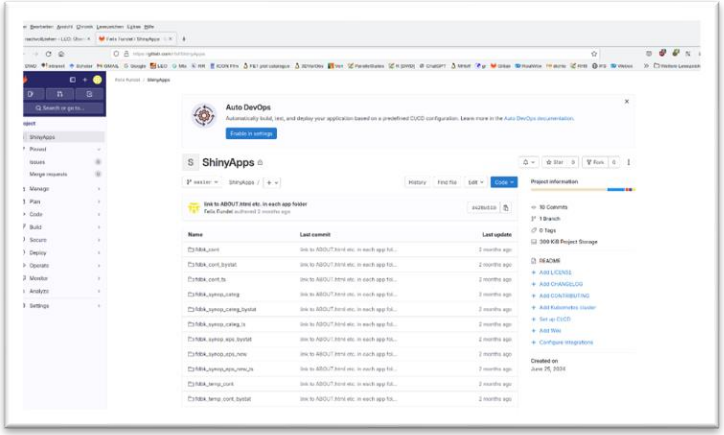

Feedback File Verification Suite at DWD
Felix Fundel
27 Juni, 2024

- Introduction
- System Requirements
  - Software
  - Hardware
- Input Data
- Workflow
  - Domain average verification
  - Station based verification
- Namelists
  - Example Namelist
  - Experiment Definition
  - Valid Time Definition
  - Forecast and Analysis Definition
  - Observation State Definition
  - Verification Period Definition
  - Variable and Threshold Selection
  - Forecast/Analysis Selection
  - Score Selection
  - Data alignment
  - Domain Stratification
  - Using Whitelists
  - Using Blacklists
  - Conditional Verification
  - Stratification by Date
  - Significance Testing
  - Hindcast & Standard Verification
  - Single Member Verification
  - Bias Corrected Observations
  - Verification Output
- Running the Verification
- Quality Control Within the Verification
  - Mimicing VERSUS Quality Check
- Visualization
  - Shiny-server applications
  - Other
- Further Tips
  - Generating Polygon Domain Tables for Countries
- Additional Tools
- Appendix
  - Scores
  - Subdomains
  - Observations

Introduction

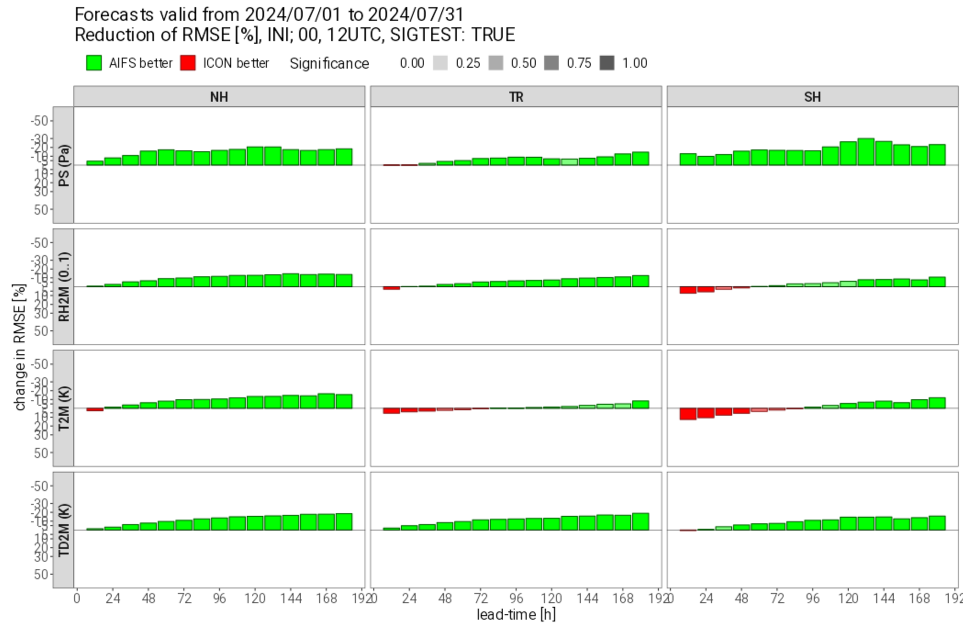
This document is meant to serve as a guideline to perform a feedback file (ff) based verification on any UNIX based system. Given that all the necessary software and scripts are available on the system, the user has to provide a set of namelists, run the verification scripts and can eventually have a look at the results through an interactive browser application. Here, the system requirements, to namelist options and tips on how to run the verification efficiently are provided.

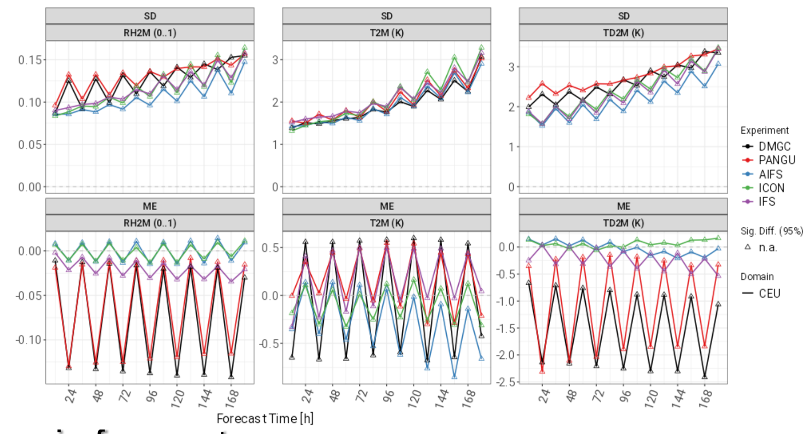
System Requirements



MEC

AIFS vs. ICON





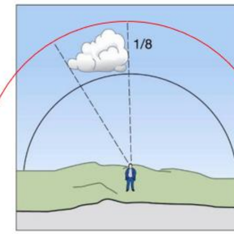
- Good performance upper air but not in very high atm. levels
- Also surface scores are good
- Some issues with humidity sensitive variables at surface
- Good performance in verification does not guarantee consistency in forecast

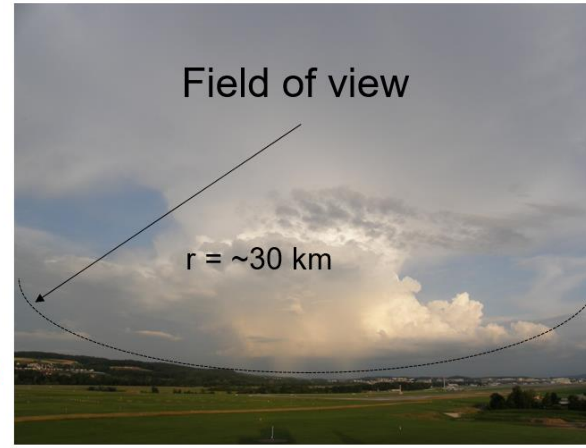
Verification Methods - High Impact Weather

Cloud and Precipitation Verification of ICON-CH Models

Pirmin.Kaufmann@MeteoSwiss.ch

Original SYNOP Eye Observations

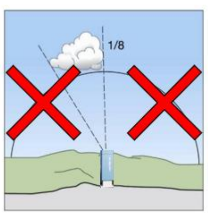


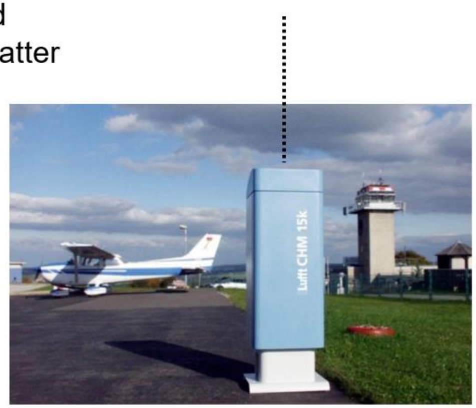


Pirmin.Kaufmann@MeteoSwiss.ch

Modern Measurement: Ceilometer

Cloud cover derived from LIDAR backscatter
→ Much narrower viewing angle



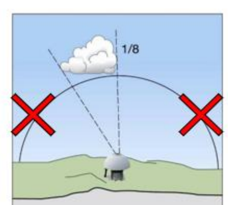


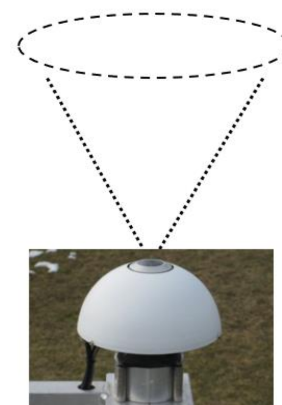
Pirmin.Kaufmann@MeteoSwiss.ch

Model	Radius approp. for SYNOP Obs.	Radius approp. for LWR Derived
ICON-CH1 (1 km grid)	26 gp ~ 27 km	4 gp ~ 4 km
ICON-CH2 (2 km grid)	13 gp ~ 27 km	2 gp ~ 4 km

Modern Measurement: Pygeometer

Cloud cover derived from incoming longwave radiation
→ Narrower viewing angle for main contribution





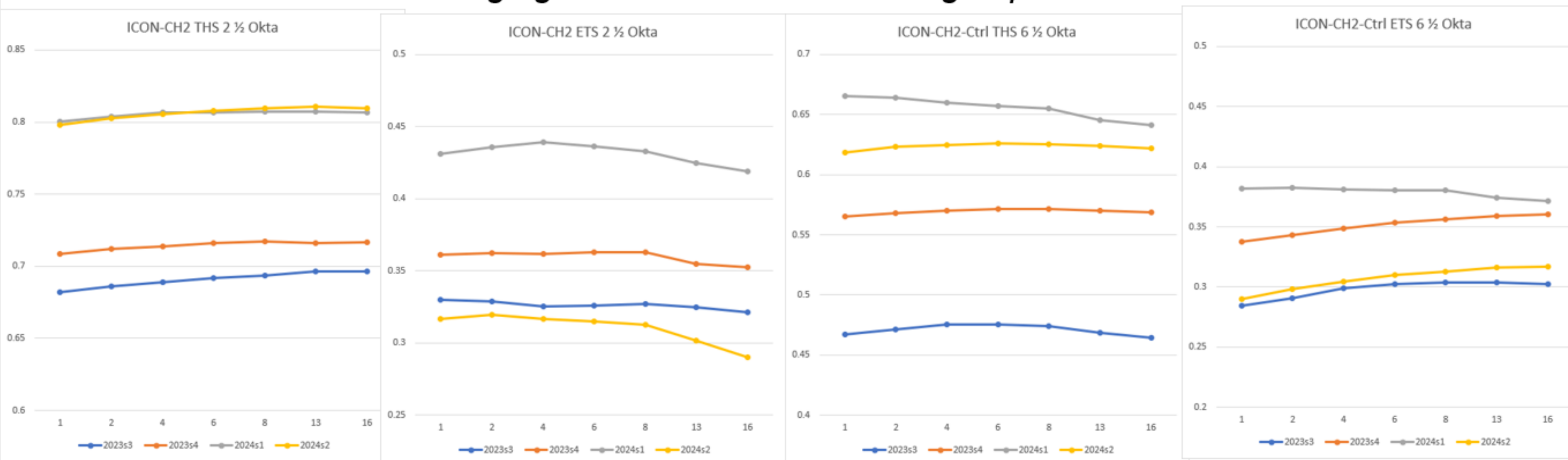
Pirmin.Kaufmann@MeteoSwiss.ch

Effect of Score, Season and Averaging window

Pirmin.Kaufmann@MeteoSwiss.ch

ICON-CH2-Ctrl, Threshold 2 ½ Okta ICON-CH2-Ctrl, Threshold 6 ½ Okta

Averaging area with radius 1 to 16 grid points



- Threat Score, depends on base frequency
  Larger averaging area → higher frequencies of events → higher THS
- ETS compensates frequency effect. Better for smaller averaging area (as opposed to THS)

- General underestimation of frequency of event
- **THS depends on base frequency, so on season**. Spring, Summer, Autumn, THS is independent of averaging area
- ETS in winter is independent of averaging area
- **optimal radius depends on season**

Model	Radius approp. for SYNOP Obs.	Radius approp. for LWR Derived
ICON-CH1 (1 km grid)	26 gp ~ 27 km	4 gp ~ 4 km
ICON-CH2 (2 km grid)	13 gp ~ 27 km	2 gp ~ 4 km

Michael Hoff

WW from Model Output

- **Why using WW: direct connection to specific weather phenomena, comparable with observed phenomena, easier to interpret than raw model output, helpful for forecaster and warnings management**
- WW calculated hourly at every model grid point via thresholding of relevant parameters (tuning necessary)
- WW contains weather of last hour
- Sometimes, ww-codes could differ from raw model output (e.g. ww=71 vs SNOW_GSP)
- Due to temporal-, spatial-, technical- constrains, not all 99 ww-codes can be calculated
- **Problems: Model precipitation (e.g. showers) mostly sums of last hour but WW observations instantaneous values leads to inconsistencies and misleading interpretation**
- Docu (engl. follows): Wetterinterpretation für die Modelle ICON, ICON-EU, ICON-D2 und COSMO-D2 (dwd.de)

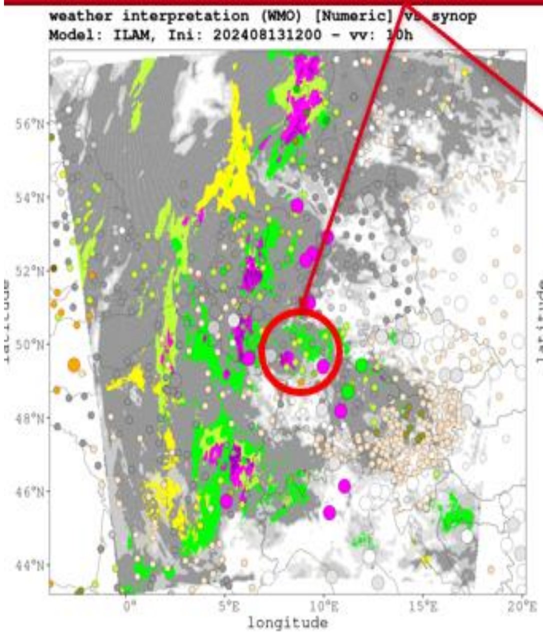
Michael Hoff
Deutscher Wetterdienst
Wetter und Klima aus einer Hand

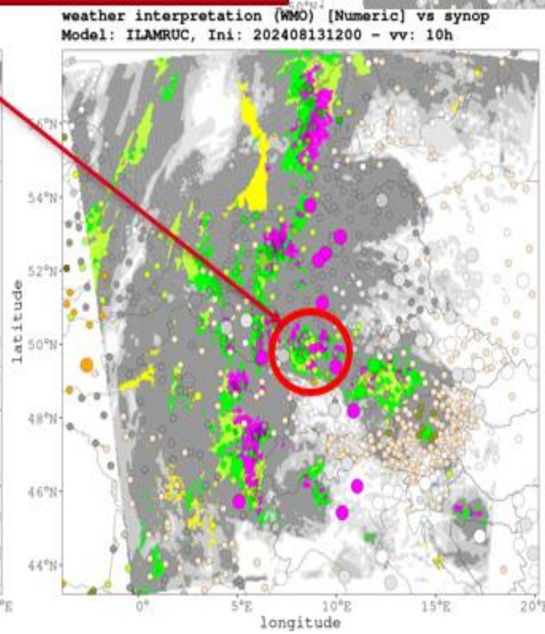
Eyeball-Verification of WW – ILAM vs. RUC

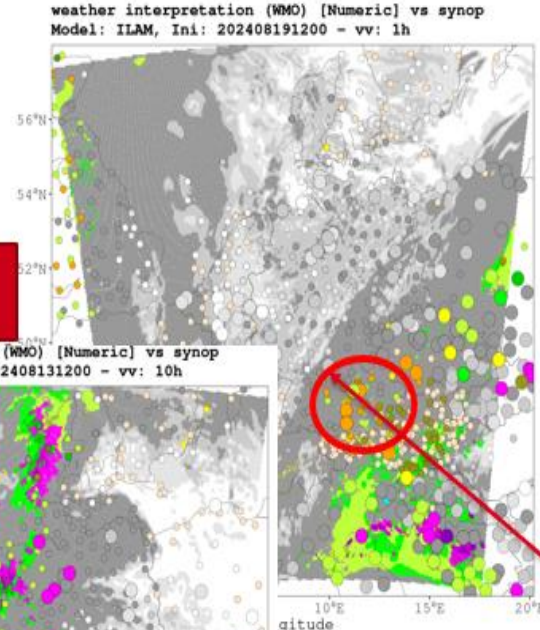
Yellow: fog
Orange: drizzle
Green: rain
Dark Green: rain Shower
Purple: thunderstorm

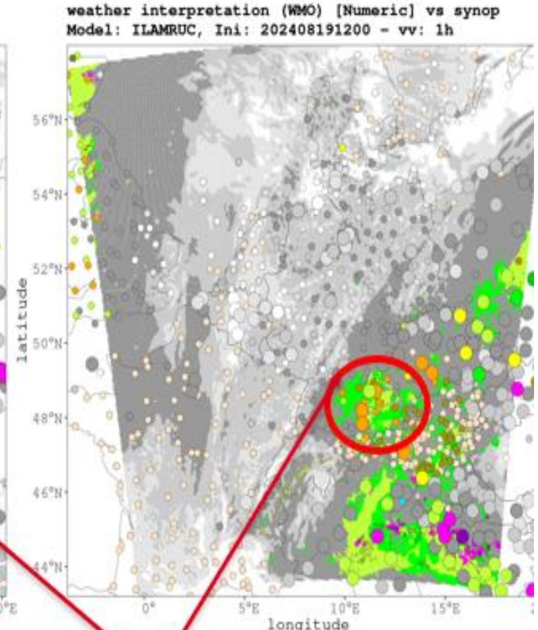
Grey: cloud cover
°: automats
o: manned stations

Intensive rain event that was almost completely missed by ILAM but partly detected by RUC







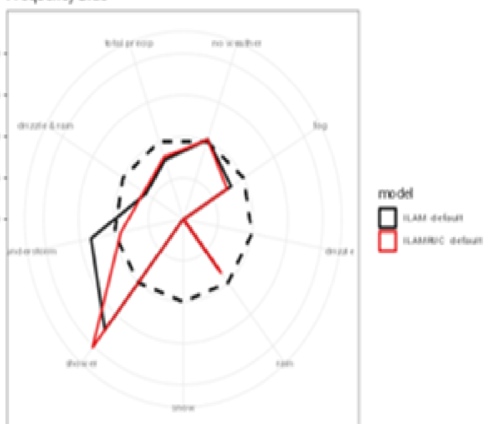


Often, RUC detects more slight rain in regions where ILAM does not detect any rain
→ therefore often better scores for RUC (at least in July/August 2024)

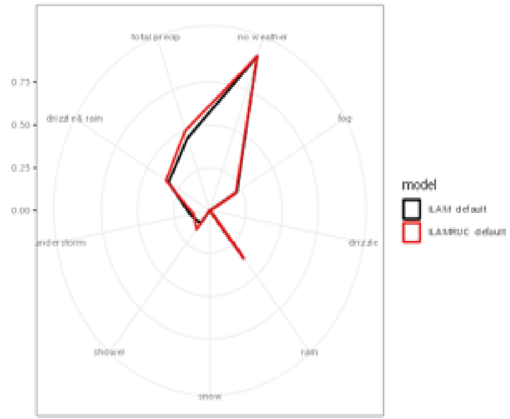
Michael Hoff
Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Radar Charts of WW-Verification – ILAM vs. RUC

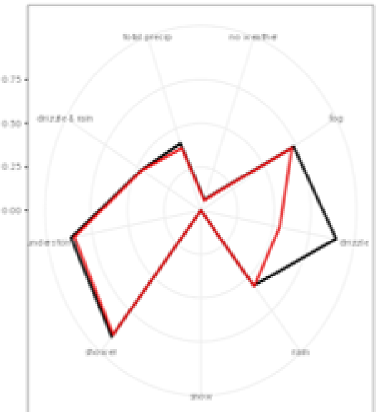
All weather verification:
Frequency Bias



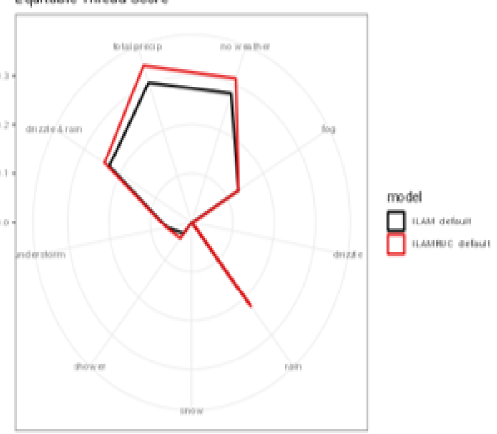
All weather verification:
Probability of Detection



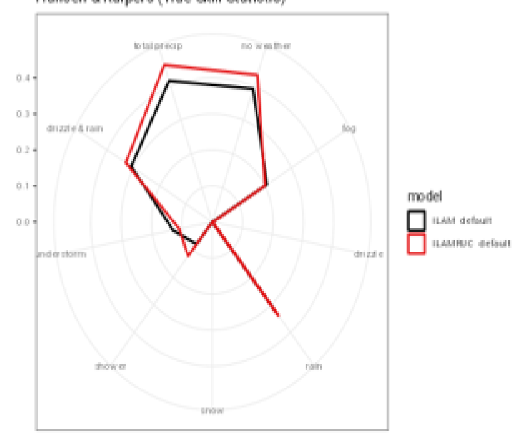
All weather verification:
False Alarm Ratio



All weather verification:
Equitable Thread Score



All weather verification:
Hansen & Kuipers (True Skill Statistic)



All weather verification:
Heidke Skill Score

August 24
00&12 UTC
+1 to +14h
Aggregated

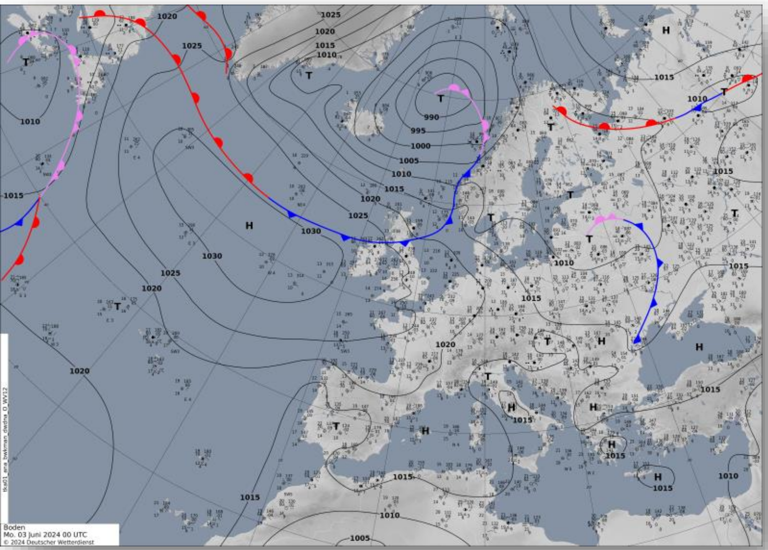
RUC has better scores for rain and weather with precipitation.

Biggest WW-Problems for high-res convection-permitting models in summer:

- too many showers
- too less drizzle
- Often shower in regions where drizzle or light rain should be
- Thunderstorms hard to tune

An AI front prediction system   -Verification Update-

Felix Fundel
Deutscher Wetterdienst
Wetter und Klima aus einer Hand



https://www.dwd.de/DE/leistungen/hobbymet_wk_europa/hobbyeuropakarten.html

Background

- Front analysis is an important product pointing to relevant weather
- Front analysis maps are hand drawn for 00 and 12 UTC
- Front forecasts for +36h, +60h, +84h and +108h from 00UTC
- Time consuming effort involving a lot of synoptic
- Since 2021 available as machine readable XML (Ninjo export)

Literature on AI for fronts

Biard and Kunkel, ASCMO, 2019
Niebler et al., WCD, 2021

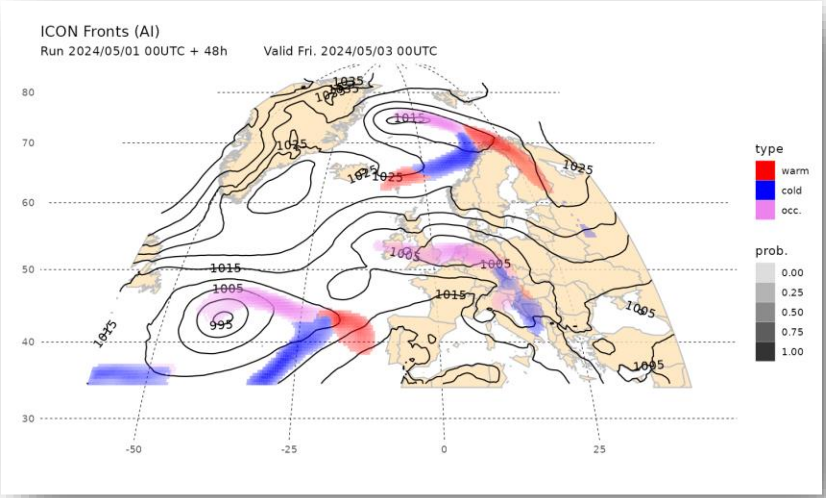
Goal

- Train an AI to draw front with NWP input
- Have a front product for every model analysis and forecast step
- Use it in verification
- Offer an AI front guidance to forecasters

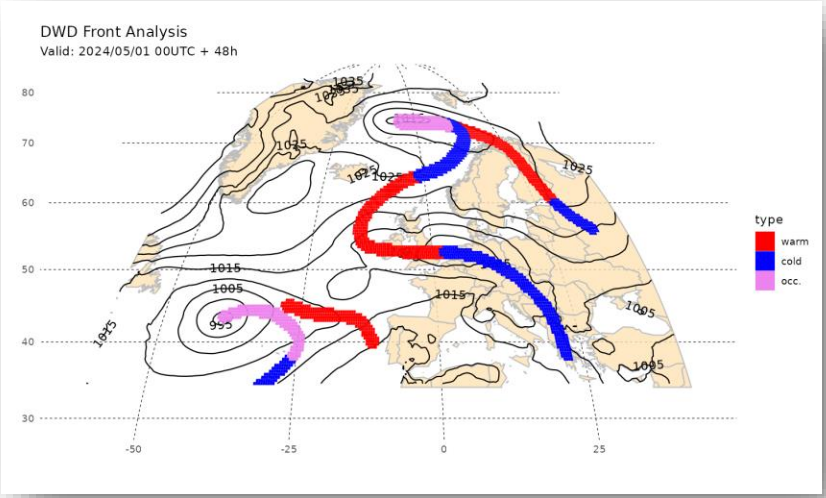
Preprocessing Model Input

- ICON global deterministic (~0.12°), vv=0
- PMSL, RH2M, T2M, U10M, V10M
- Re-gridded to ~0.4° (256x128 grid points)
- Each variable normalized

AI 48h front forecasts



Human 48h front forecasts



- May 2024 AI and human front forecasts
- Realistic front classification and position
- AI fronts come with probability, thus broader, reflecting the model uncertainty

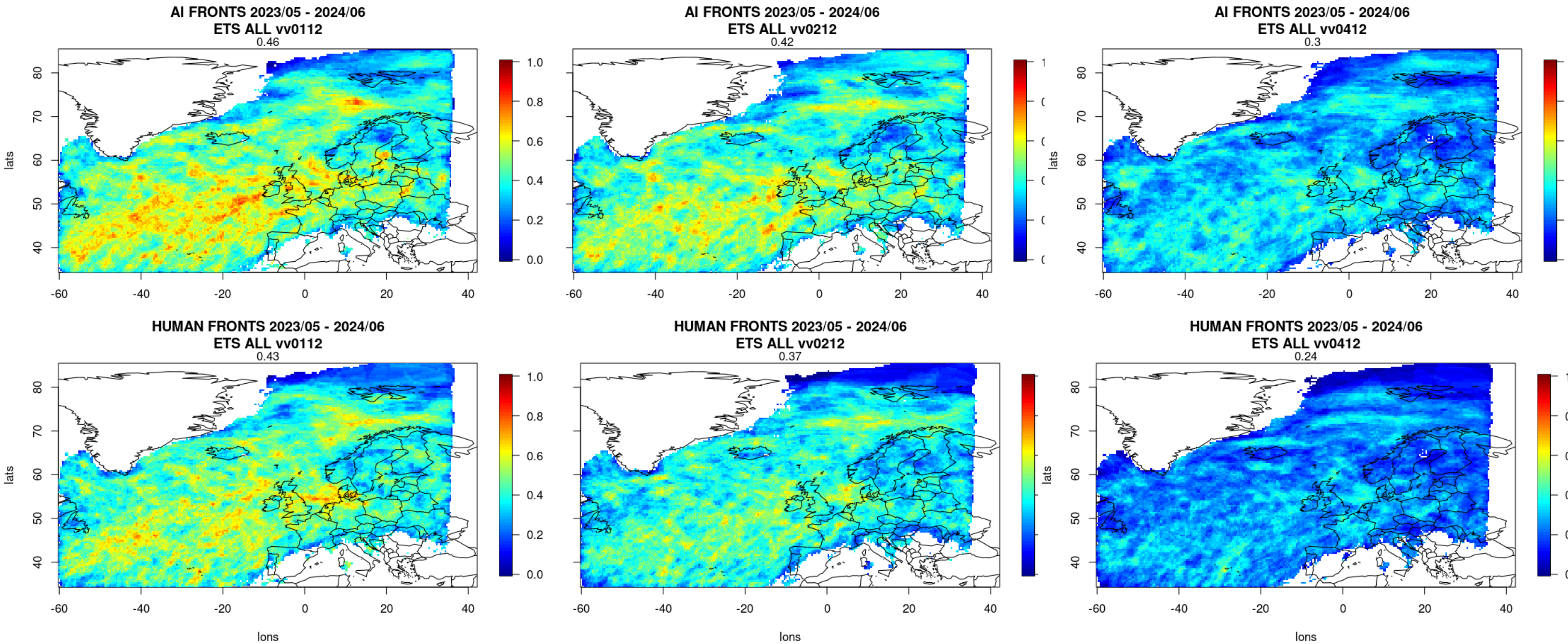
Evaluation (against human analysis)

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

+1.5 days | +2.5 days | +4.5 days

AI

Human



AI forecasts outperform forecaster forecasts, reasons:

- **AI not subject to personal preferences, time-pressure, previously made analysis, i.e. consistency**
- **AI and forecaster input is the same (ICON model), no advantage for forecasters**

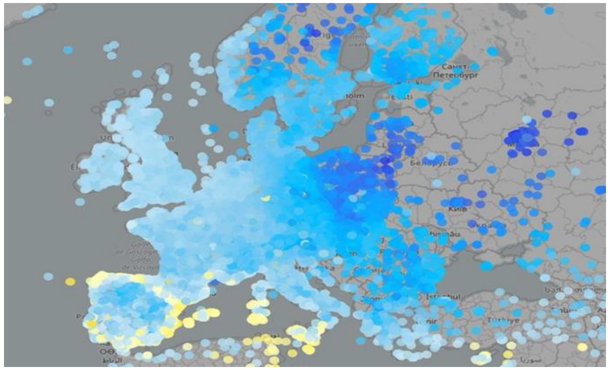
Projects - Collaborations

Priority Task: EPOCS (Evaluate Personal Weather Station and Opportunistic Sensor Data CrowdSourcing)

PL: Joanna Linkowska, IMGW-IB

The aim of PT EPOCS was to assess the use of weather data from Personal Weather Stations (PWS) and other Opportunistic Sensors (OS).

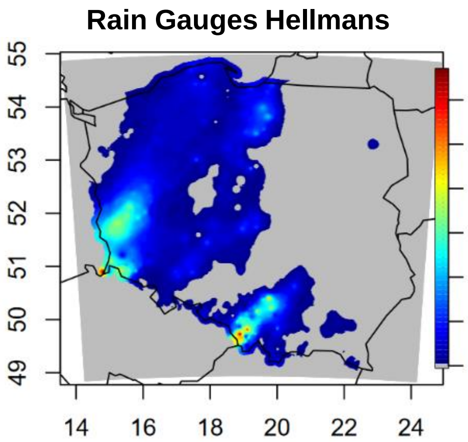
1. PWS databases survey and exploitation
2. QC algorithms for precipitation (RainGaugeQC algorithms , open-source software package TITAN)
3. Analysis of PWS based gridded rainfall products (Processing different rainfall data sources, Reliability analysis of a gridded high-resolution estimates of precipitation)

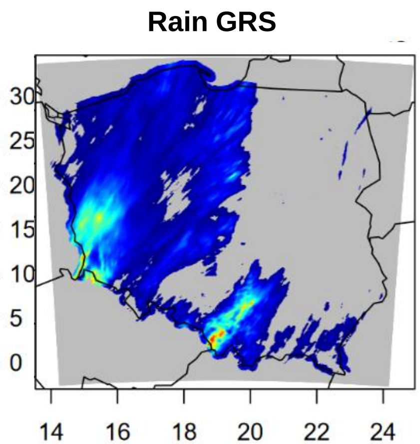


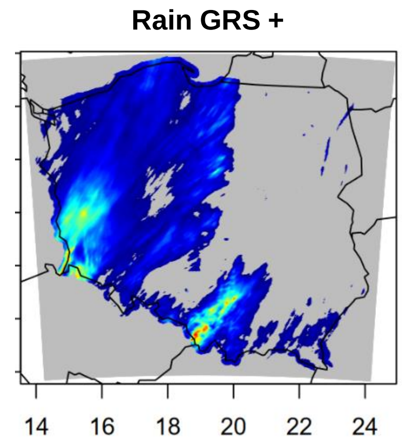
IMGW-PIB: Joanna Linkowska, Jan Szturc, Anna Jurczyk, Katarzyna Ośródka, Marcin Grzelczyk, +Radosław Droździoł
CIMA: Massimo Milelli, Elena Obert, Umberto Pellegrini
CNMCA: Francesco Sudati

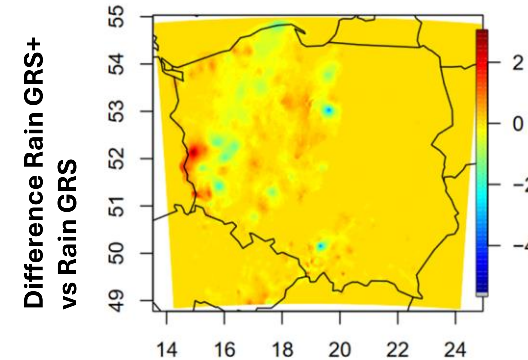
3.2 Reliability analysis of a gridded RainGRS+ high-resolution estimates of precipitation - case study 20230919-20231001- 24h precipitation (Joanna Linkowska)

METEO IMGW-PIB meteo.imgw.pl









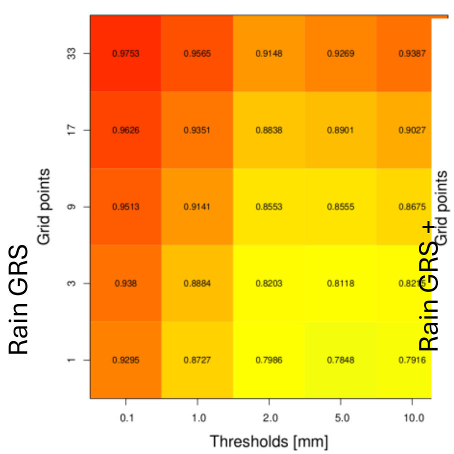
Rain GRS v s Hellmann Rain Gauges

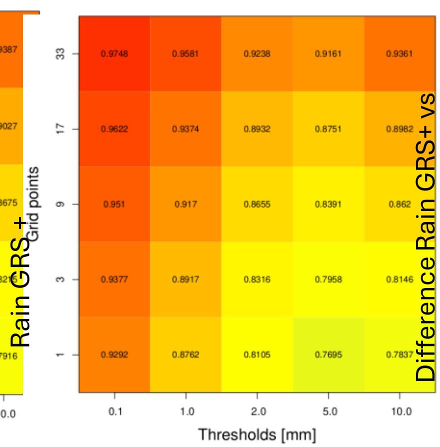
Th [mm]	0.1	1.0	2.0	5.0	10.0
POD	0.9108	0.9265	0.8853	0.7878	0.7504
FAR	0.1055	0.1150	0.1288	0.1523	0.2944
BIAS	1.0182	1.0470	1.0162	0.9293	1.0635
ETS	0.7774	0.7990	0.7572	0.6735	0.5651

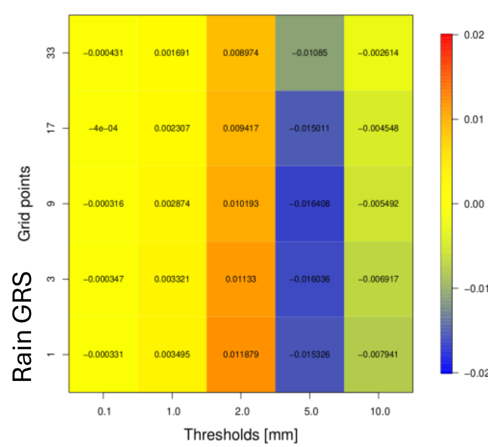
Rain GRS + v s Hellmann Rain Gauges

	0.1	1.0	2.0	5.0	10.0
POD	0.9116	0.9283	0.8908	0.8043	0.7651
FAR	0.1042	0.1115	0.1294	0.1584	0.3048
BIAS	1.0176	1.0447	1.0233	0.9557	1.1006
ETS	0.7796	0.8042	0.7612	0.6821	0.5664

- ❑ Slightly better POD and ETS for RainGRS + for all the thresholds,
- ❑ FAR slightly better only for small precipitation.
- ❑ BIAS better for all the thresholds except 2.0mm and 10.0mm.





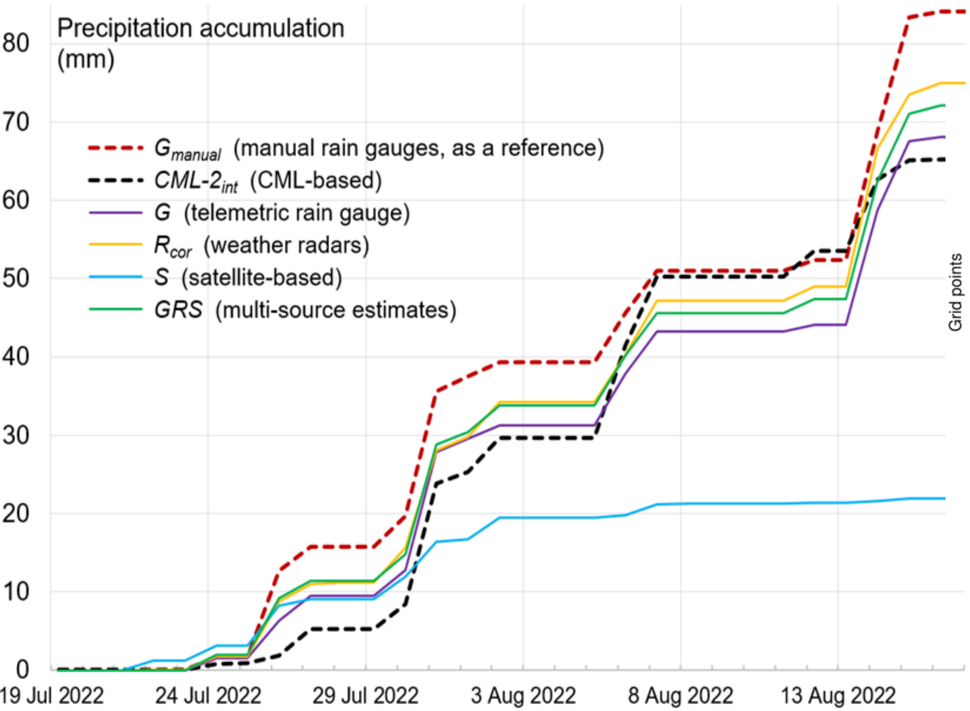


- ❑ For all thresholds & spatial scales FSS > 0.7
- ❑ Very high values of FSS for smaller thresholds and smaller & larger spatial scales
- ❑ Difference of FSS for Rain GRS & Rain GRS + are neglectable

3.2 Reliability analysis of a gridded RainGRS+ high-resolution estimates of precipitation - CML based precipitation case study (Joanna Linkowska)

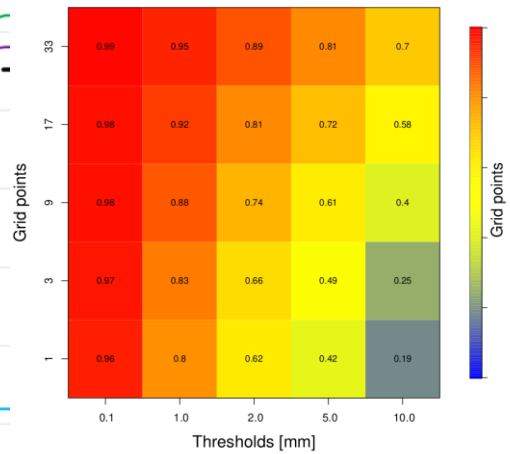
METEO IMGW-PIB meteo.imgw.pl

24h precipitation (mm) Opole station

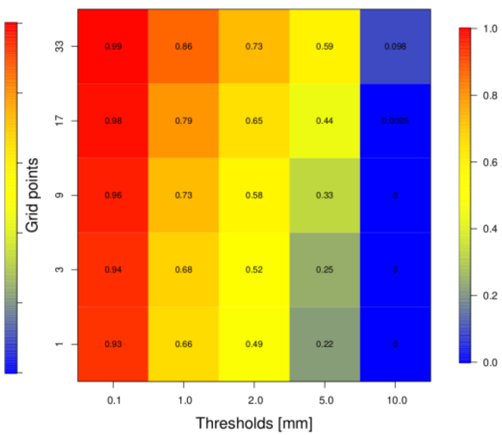


Correct detection of the precipitation using CML-based estimates (CML-2) with respect to the manual rain gauge, but not quite satisfactory estimation of the precipitation amount.

FFS: CML vs RainGRS



CML vs Hellman Rain Gauges



- ❑ Higher FSS values for CML compared to Rain GRS than against Rain Gauges.
- ❑ Very high or high FSS for smaller thresholds and small and larger spatial scales.
- ❑ Better scores for RainGRS than Rain Gages for higher thresholds and larger spatial scales

❑ The spatial precipitation distribution captured relatively well

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

PL: Martyna Zosicz, MGW-PIB

Duration 2 years, to initiate in early 2025

Objectives of the project:

- Building a useful COSMO database with PWS data.
- Application and testing of the developed RainGaugeQC and Titan - Quality Control (QC) software packages.
- Evaluation of the quality of various physical parameters, including precipitation, temperature, and humidity.
- Application of products developed from PWS data for model assimilation and numerical forecast verification.

IMGW-PIB: Martyna Zosicz, Jan Szturc, Anna Jurczyk, Katarzyna Ośródka, Magdalena Pasierb, Marcin Grzelczyk, Radosław Drożdżoł, Bartłomiej Sobczyk, Adam Jaczewski, Joanna Linkowska (?)
HNMS: Flora Gofa, Dimitra Boucouvala
CIMA: Massimo Milelli, Elena Oberto, Francesco Uboldi
ARPA Piemonte: Valeria Garbero
Politecnico di Torino: Tanguy Houget
CNMCA: Francesco Sudati

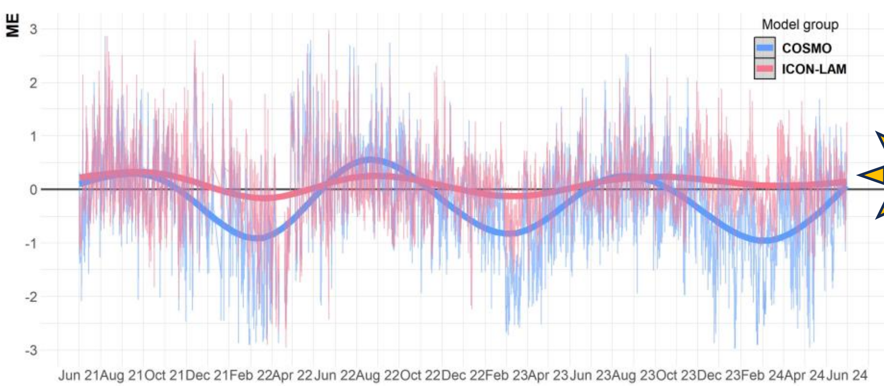
2mT @12+24UTC: all COSMO/ICON ComA2. 2021-2024

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

Time series: 2mT ME | Region: ComA2

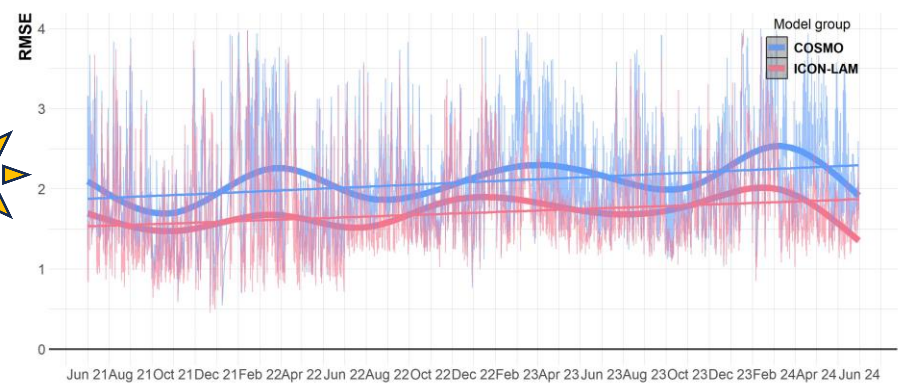
Forecasting Period: 06/2021-05/2024 | 12:00 UTC



RMSE

Time series: 2mT RMSE | Region: ComA2

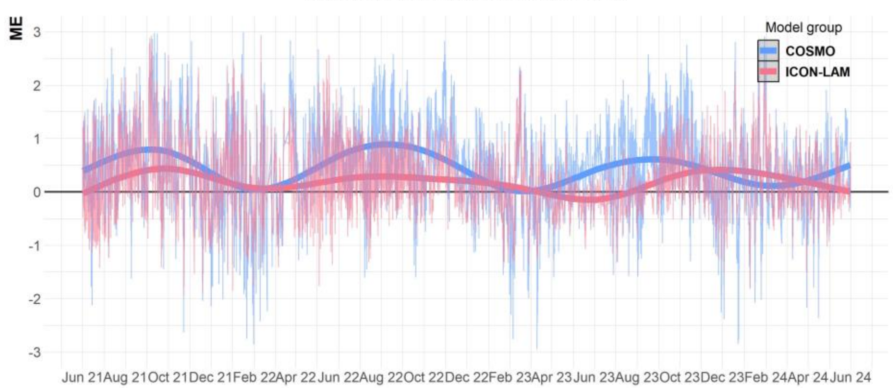
Forecasting Period: 06/2021-05/2024 | 12:00 UTC



Mean Error

Time series: 2mT ME | Region: ComA2

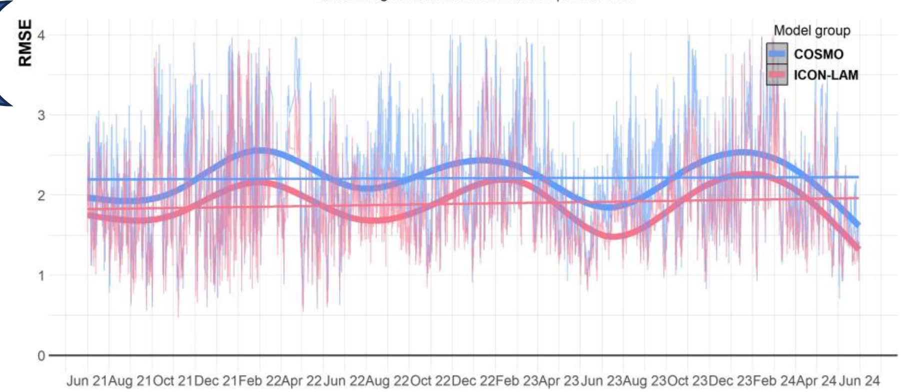
Forecasting Period: 06/2021-05/2024 | 24:00 UTC



RMSE

Time series: 2mT RMSE | Region: ComA2

Forecasting Period: 06/2021-05/2024 | 24:00 UTC



More than 3year operational experience: ICON is a good model, shares some common flaws and strengths as COSMO, susceptible of improvement

Performance Trends derived from Common Plots

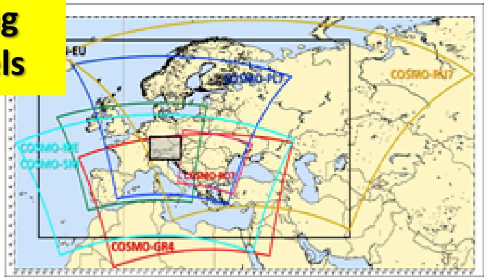
CONSORTIUM FOR SMALL SCALE MODELING COSMO

Common Plots: redefined areas

Drivng Models

ComA-1

Specs



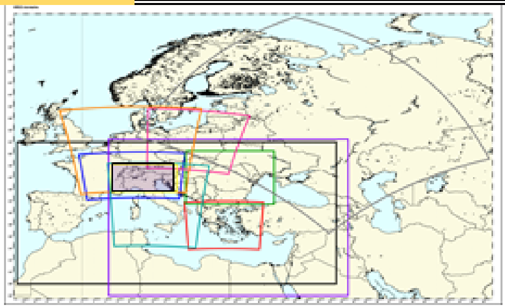
Forecast run: 00UTC
Forecast Horizon: 48h
Seasons: JJA22, SON22, DJF23, MAM23

Models

Global: ICON, IFS
LAMS: DWD: ICON-EU, COMET: COSMO-ME
IMGW-PIB: COSMO-PL7

FINE — ComA-2

Specs



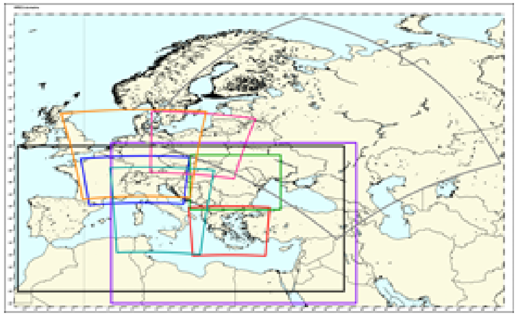
Forecast run: 00UTC
Forecast Horizon: 48h
Seasons: JJA22, SON22, DJF23, MAM23
Area: 43.5/5.0/48.2/16.0

Models

Driving models: ICON-EU, IFS, ICON
LAMs: DWD: ICON-D2, MCH: COSMO-1E (control), COSMO-2E, HNMS: ICON-GR
COMET: COSMO-I2, ICON-I2, ARPA-E: COSMO-2I
IMS: ICON_IL2p5

MIX — NoComA (National Domains)

Specs



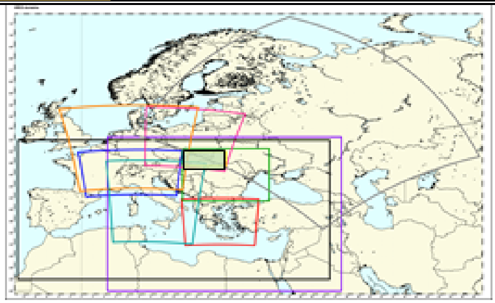
Area: national domains
Forecast Horizon: variable
Seasons: JJA22, SON22, DJF23, MAM23

Models

COSMO and ICON-LAM
DWD, MCH, COMET, HNMS, IMGW-PIB, NMA, RHM, IMS, ARPA-E

FINE — ComA-3

Specs



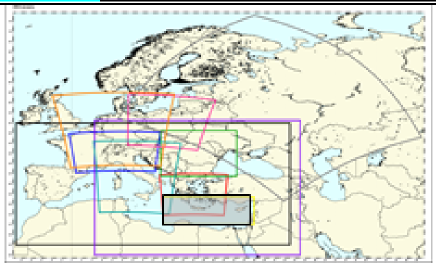
Forecast run: 00UTC
Forecast Horizon: 48h
Seasons: JJA22, SON22, DJF23, MAM23
Area: 47.5/17.7/50.0/25.0

Models

Driving models: ICON-EU, IFS, ICON
LAMs: HNMS: ICONGR2.5, NMA: COSMO-NMA/ICON-NMA. IMGW-PIB: ICON-PL, IMS: ICON–IL2.5

MIX — ComA-TCC

Specs



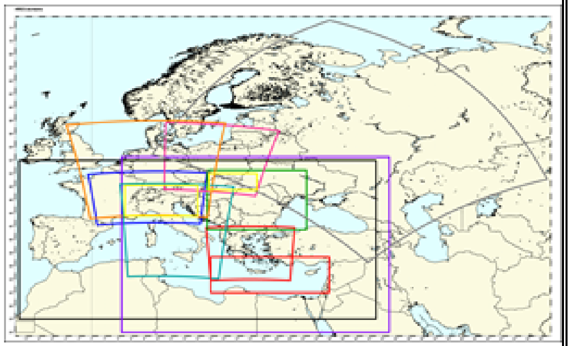
Forecast run: 00UTC
Forecast Horizon: 48h
Period: June2022, Dec2022, Apr2023
Area: 31.0/18.0/37.4/32.0

Models

Driving models: IFS, ICON
LAMs: HNMS: ICONGR2.5, IMS: ICON_IL2p5, ICON-EU

Optional — ComA-OnDemand

Specs



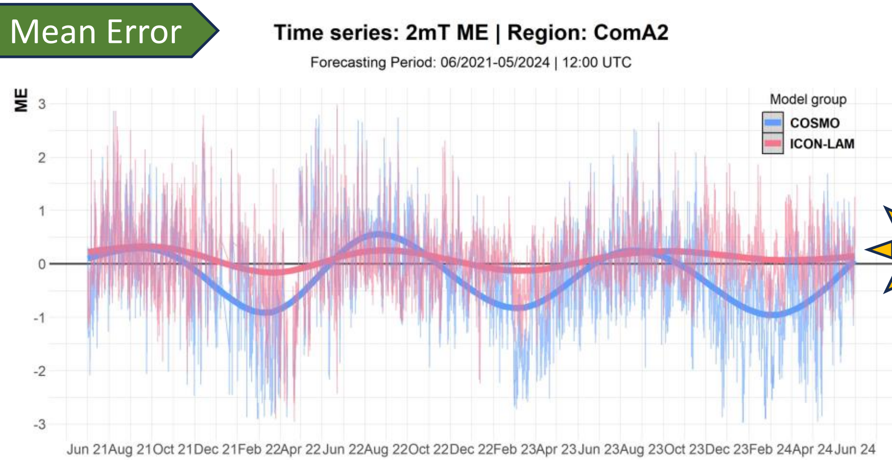
Models

Area and LAMs, Specs based on specific experiment

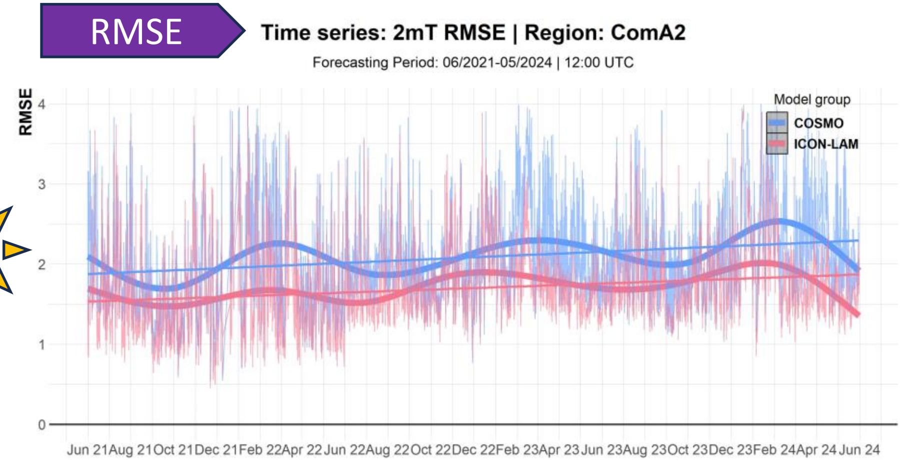
2mT @12+24UTC: all COSMO/ICON ComA2. 2021-2024

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.

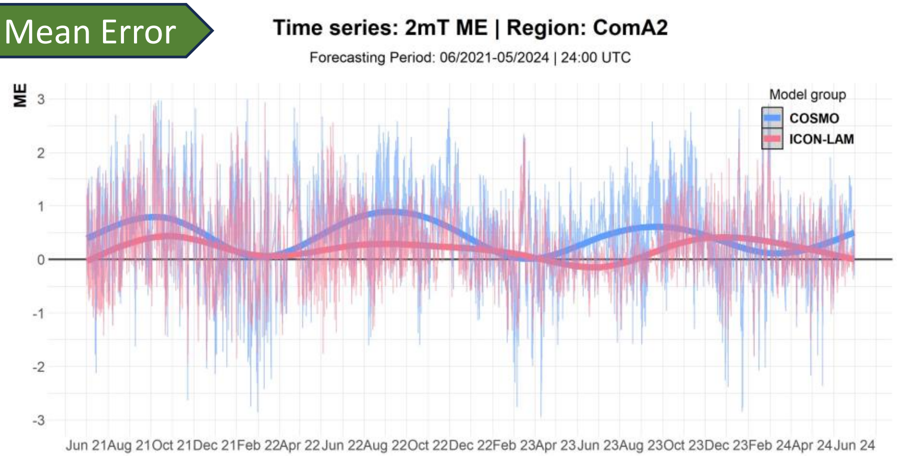
Mean Error



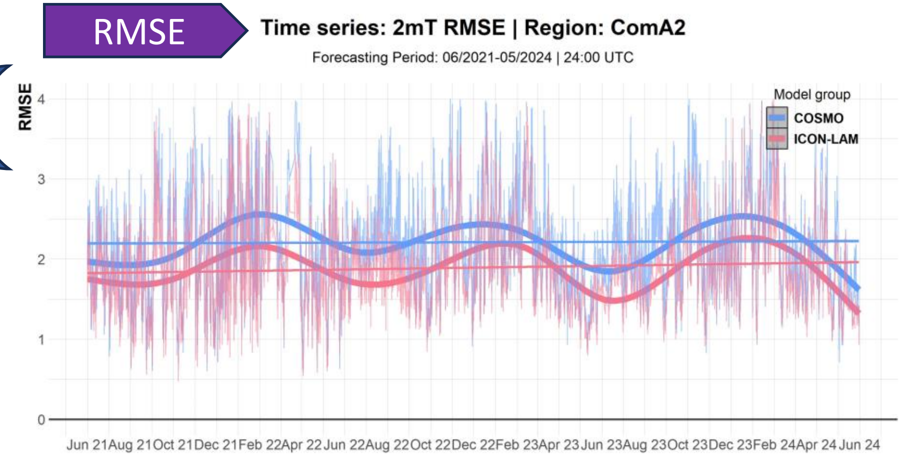
RMSE



Mean Error



RMSE

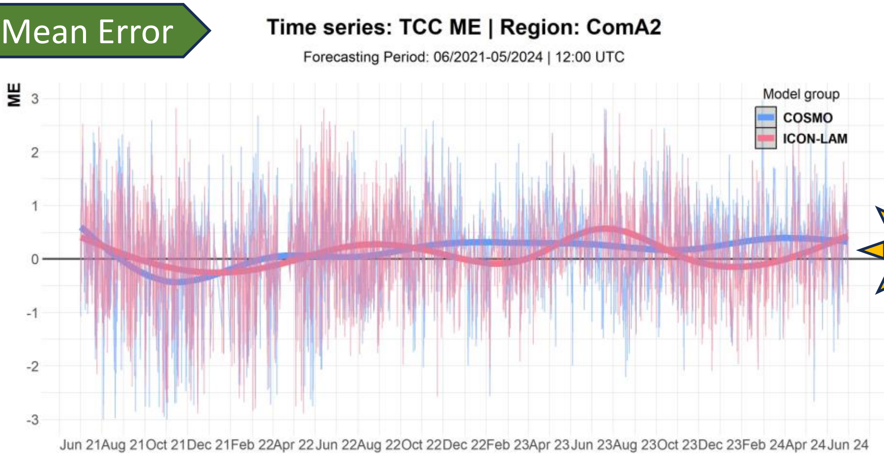


More than 3year operational experience: ICON is a good model, shares some common flaws and strengths as COSMO, susceptible of improvement

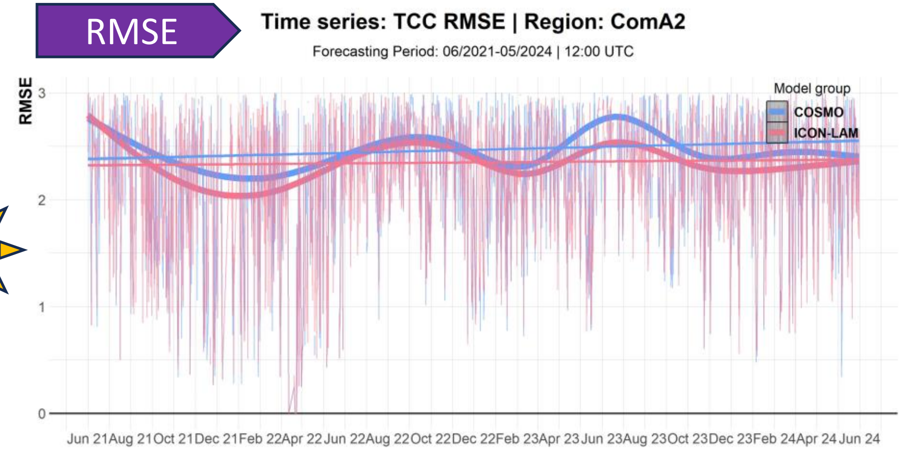
TCC @12+24UTC: all COSMO/ICON ComA2, 2021-2024

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.

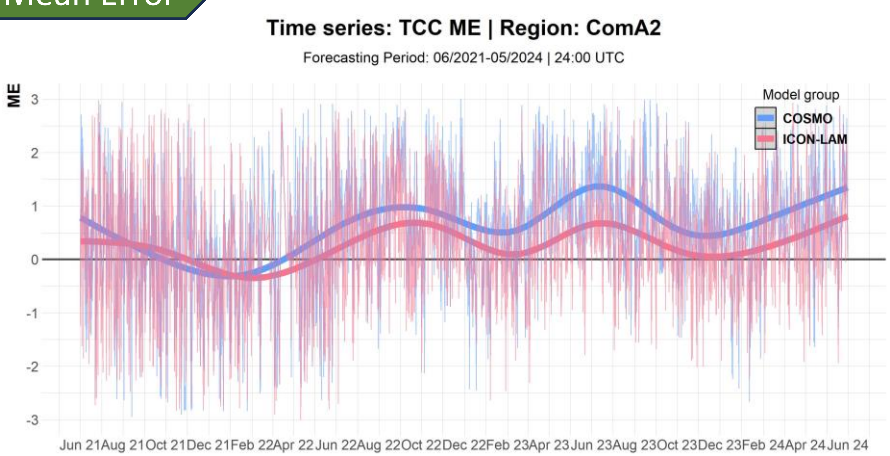
Mean Error



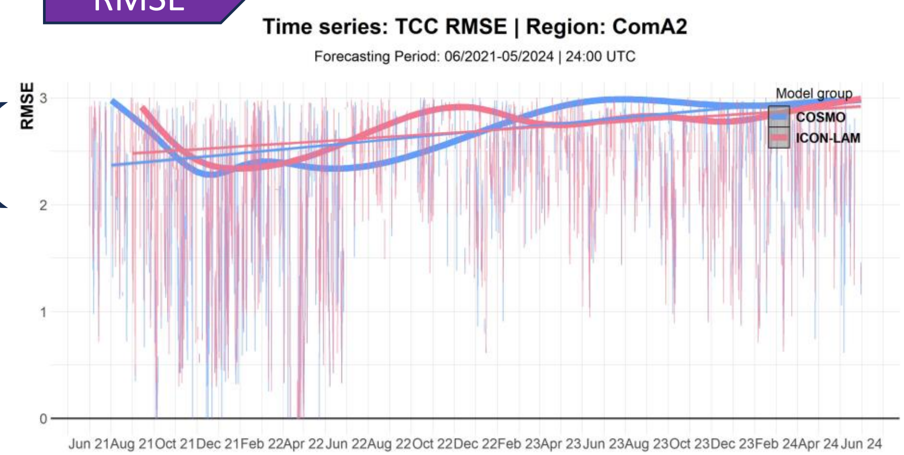
RMSE



Mean Error

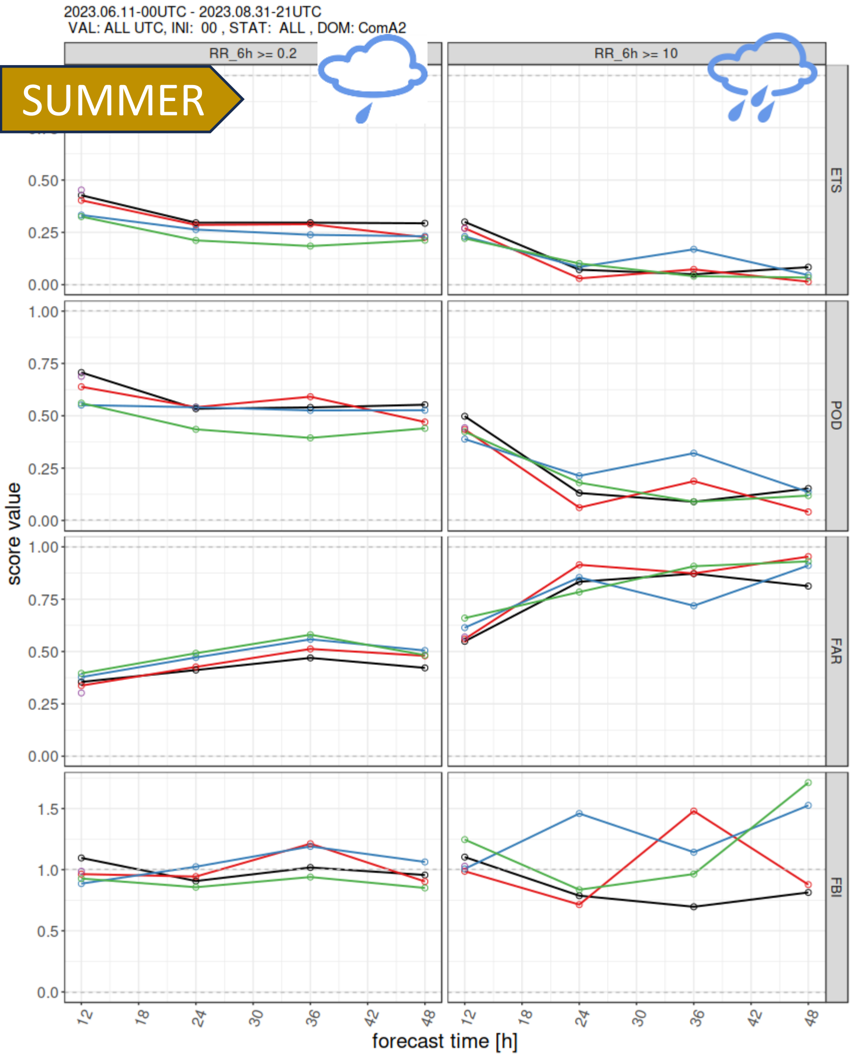


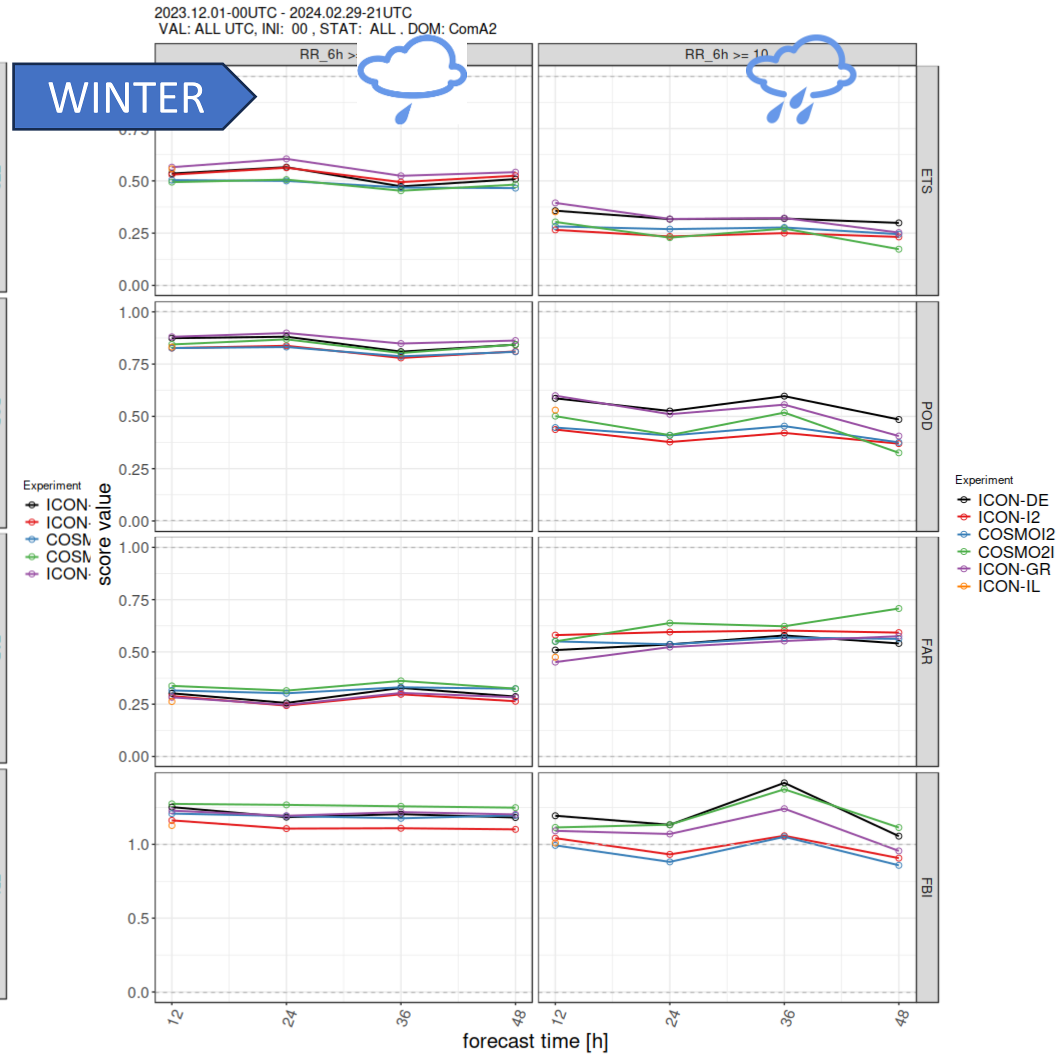
RMSE



Ambiguous performance in RMSE, both models tend to overestimate during warm months esp. @night hours (model dependant). Varying performance among installations.

COSMO/ICON ComA2 6h Precipitation





Success ratio (1 - FAR)

➢ Clear differences in performance

Seasonal average: ETS, POD (also FAR) are higher for **ICON** but with a tendency to **underestimate small precipitation amounts**

HIW cases: ICON exaggerates precipitation peaks more than COSMO

WG:V/A activities contributors

F. Fundel, J. Linkowska, N. Vela, S. Gabrian, S. Dinicila, M.S.Tesini, F. Sudati, F. Batignani, D. Boucouvala, P. Kaufmann, M. Zosicz, M. Grzelczyk, P. Khain, B. Maco, F.Gofa,….

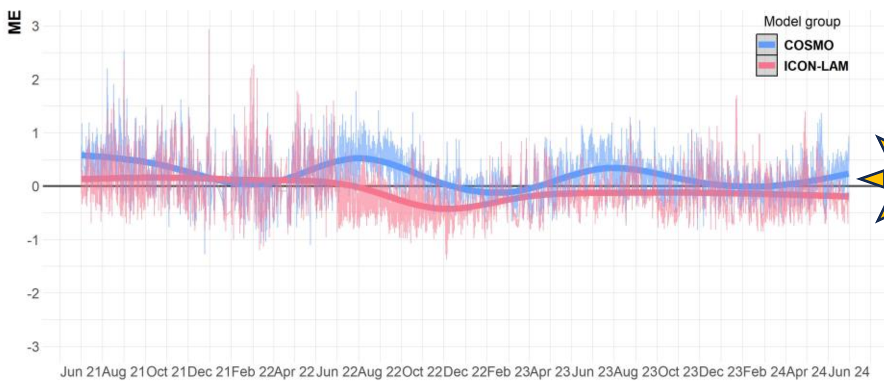
DĚKUJI VÁM!

WindSp @12+24UTC: all COSMO/ICON ComA2, 2021-2024

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.

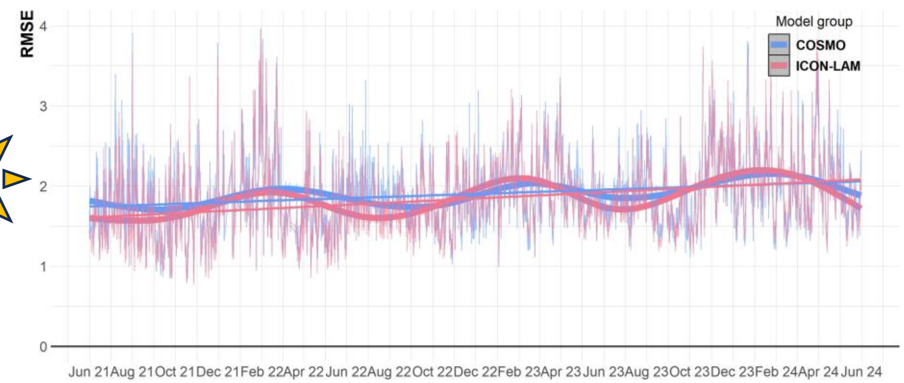
Mean Error

Time series: WindSp ME | Region: ComA2
Forecasting Period: 06/2021-05/2024 | 12:00 UTC



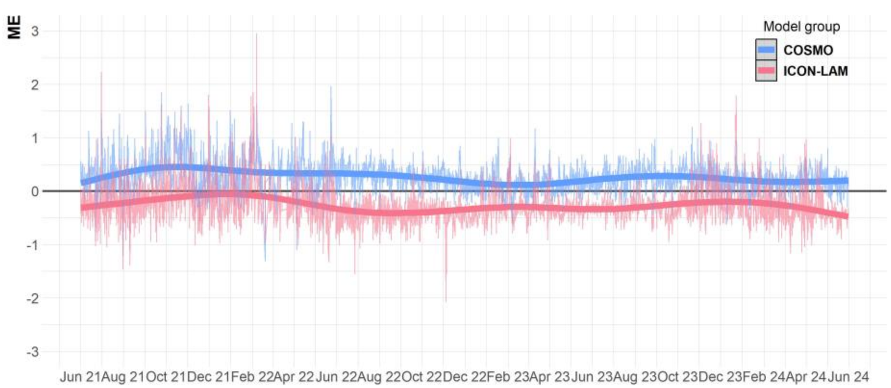
RMSE

Time series: WindSp RMSE | Region: ComA2
Forecasting Period: 06/2021-05/2024 | 12:00 UTC



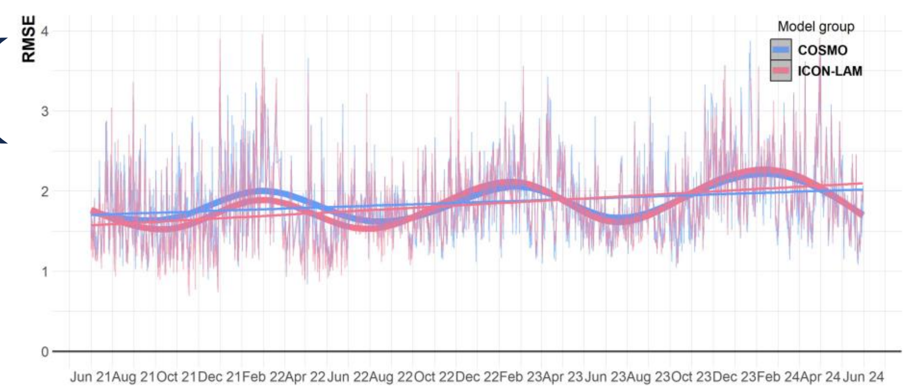
Mean Error

Time series: WindSp ME | Region: ComA2
Forecasting Period: 06/2021-05/2024 | 24:00 UTC



RMSE

Time series: WindSp RMSE | Region: ComA2
Forecasting Period: 06/2021-05/2024 | 24:00 UTC



Smaller change in performance in Wind speed, tendency to underestimate mainly at night