Comparing Models: Severe Weather Verification at ZAMG

Phillip Scheffknecht

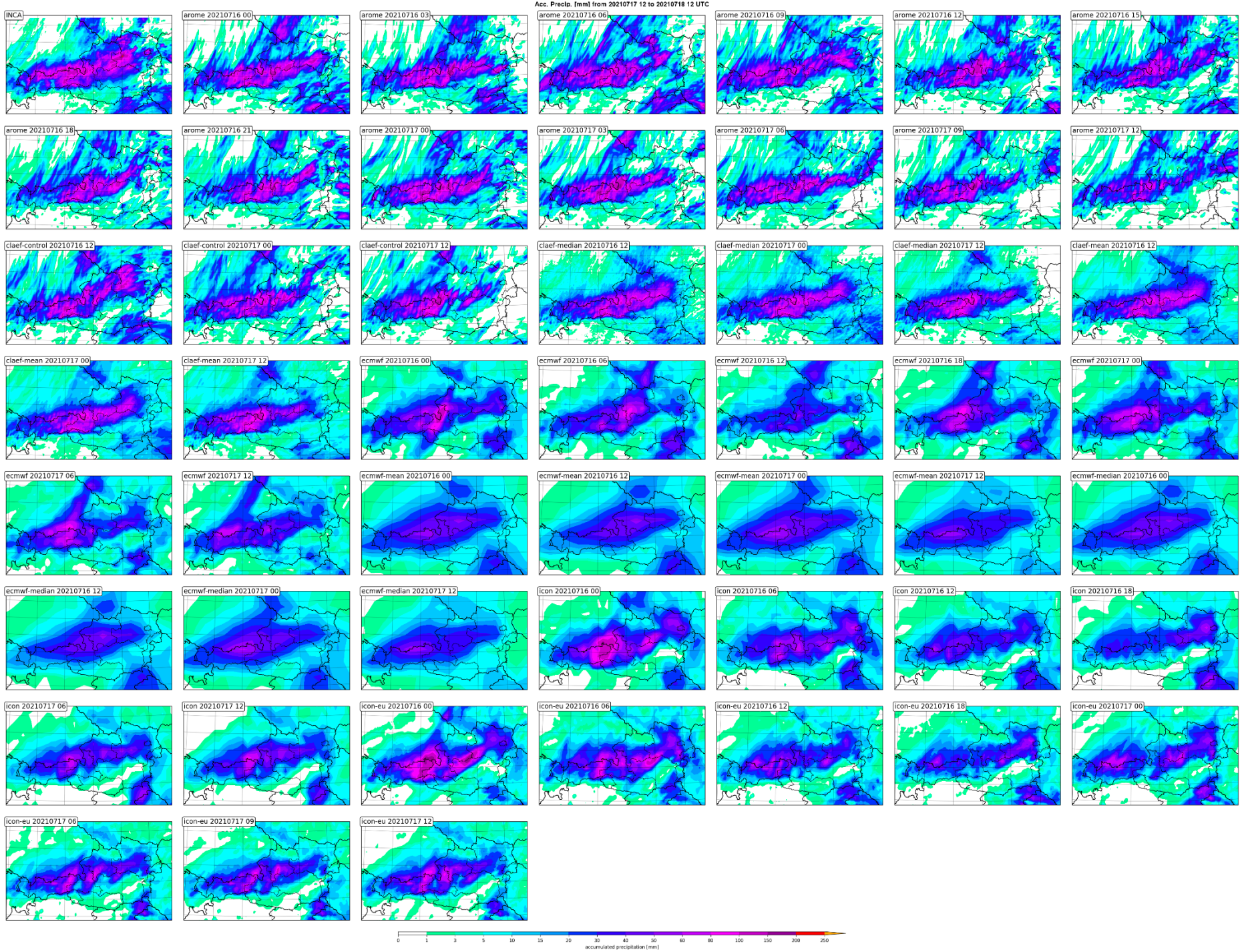
ZAMG
Zentralanstalt für
Meteorologie und
Geodynamik

Overview

What is the challenge?

Upcoming Parameters

A Look at the Severe Storm of 24 June 2021



The Challenge

This is **only a part of the models** that are available at all times.

Forecasters can have their own preferences, but in case of a severe event, **we must compare and investigate**.

Where do we **begin???**

Verification – the Eternal Struggle

- We wanted a **tool for quick model comparisons** after **warning events and fore case studies** at ZAMG

- **Basic requirements and context:**
  - **Compare forecasts,** focus is on events
  - There is **a low enough number of forecasts to look at each** at least briefly
  - Give a **quick estimate**, which **might** be „the best“ – even if it‘s crude
  - Focus on **visual presentation** of the results
  - If possible **aid** the expert in **presenting** the results they find

Verification – the Eternal Struggle

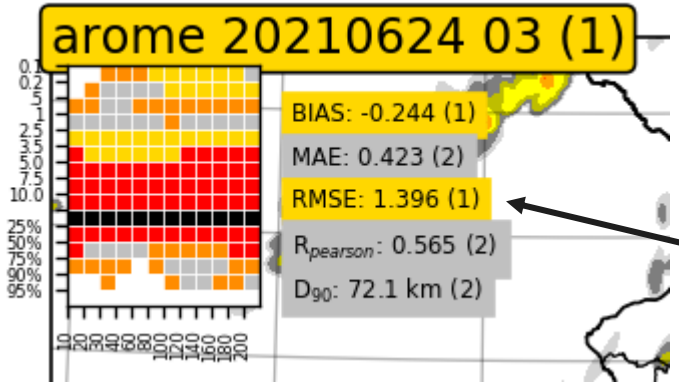
- We wanted a **tool for quick model comparisons** after **warning events and fore case studies** at ZAMG

- **Basic requirements and context:**
  - **Compare forecasts,** focus is on case studies
  - **Present results in a consistent way**
  - There is **a low enough number of forecasts to look at each** at least briefly
  - **Visualize all fields and show them together**
  - Give a **quick overview**, which might be „the best“ – even if it‘s crude
  - **Ranking suggestions – even if not fully accurate, the visualized fields will show it**
  - Focus on **visual presentation** of the results
  - **Add scores to the presentation, so they are not hidden elsewehere***
  - If possible **aid** the expert in **presenting** the results they find
  - **Make presentable graphics**

***unless we want to hide them**

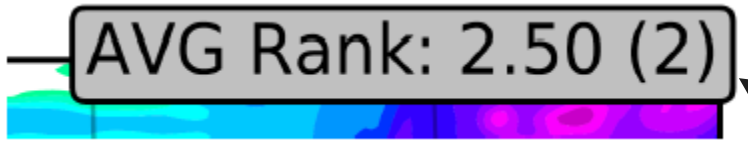
ZAMG
Zentralanstalt für Meteorologie und Geodynamik

Scoring and Ranking in Panels (1)



Scores (rank among sample)

- BIAS
- Mean Absolute Error
- Root Mean Square Error
- Pearson Correlation
- Displacement of the 90th percentile of precipitation



Average Rank (rank)

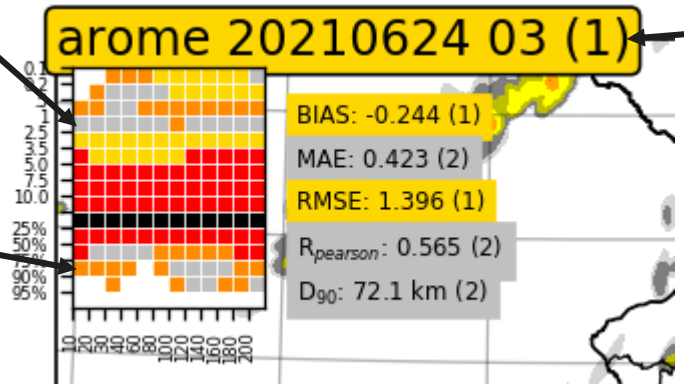
Averaged rank from **BIAS, MAE, RMSE, and Pearson Correlation**

Experimental ranking, does not always work well

Scoring and Ranking in Panels (2)

FSS for absolute thresholds in mm during the verification period

FSS for the percentiles, i.e. unbiased precipitation field, only the location is taken into account



Suggested rank based on FSS from **absolute** thresholds

$$RankScore = \sum_{i=0}^{i=N_w} \sum_{j=0}^{j=N_{th}} \frac{1}{2 + Rank}$$

Color	Meaning
	Perfect score of 1.0
	Rank 1 (if none are perfect)
	Rank 2 (if rank 2 is not perfect)
	Rank 3 (if rank 3 is not perfect)
	Not in top 3 but **above useful and skillful threshold**
	Below useful and skillful threshold or part of the verification domain is outside the model domain
	Threshold is above observed value (FSS produces NaN)

The final ranking is determined by the RankScore

This ranking is experimental, but was found to agree relatively well with what experts .

Ranking is not comparable between different sets of panels, it is **valid only within the shown sample!**

Overview

What is the challenge?

Upcoming Parameters

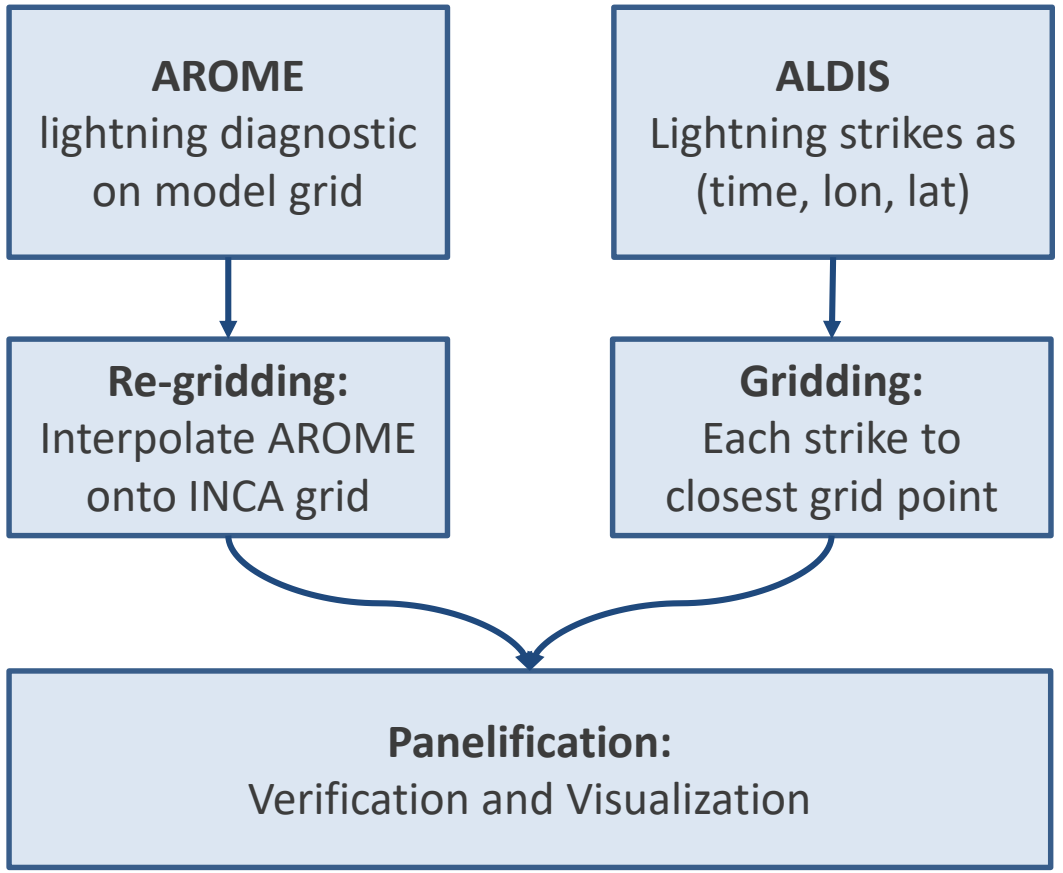
A Look at the Severe Storm of 24 June 2021

New Variable 1: Lightning

- **OBS**: ALDIS lightning strikes from the ZAMG data base
- **Model**: lightning diagnostic by McCaul et al. (2009) from AROME simulations

- Advantages:
  - Easy to detect
  - Reasonably well localized
  - Easy to count, good quantitative data
  - Great for exact location of heavy convection (more strongly linked to the column of rising warm air than precipitation)

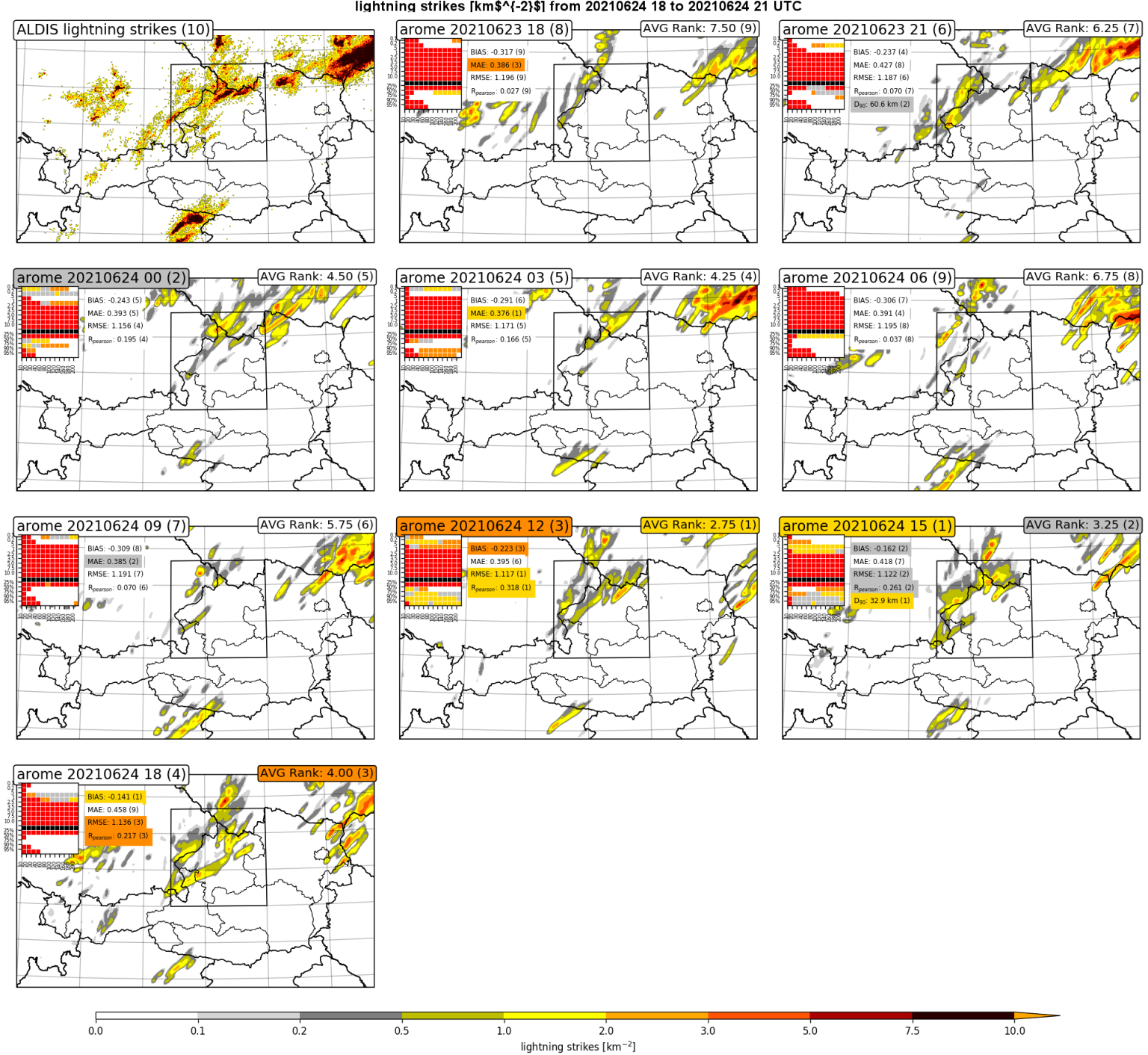
- Caveats:
  - The Diagnostic itself is tuned from Observations



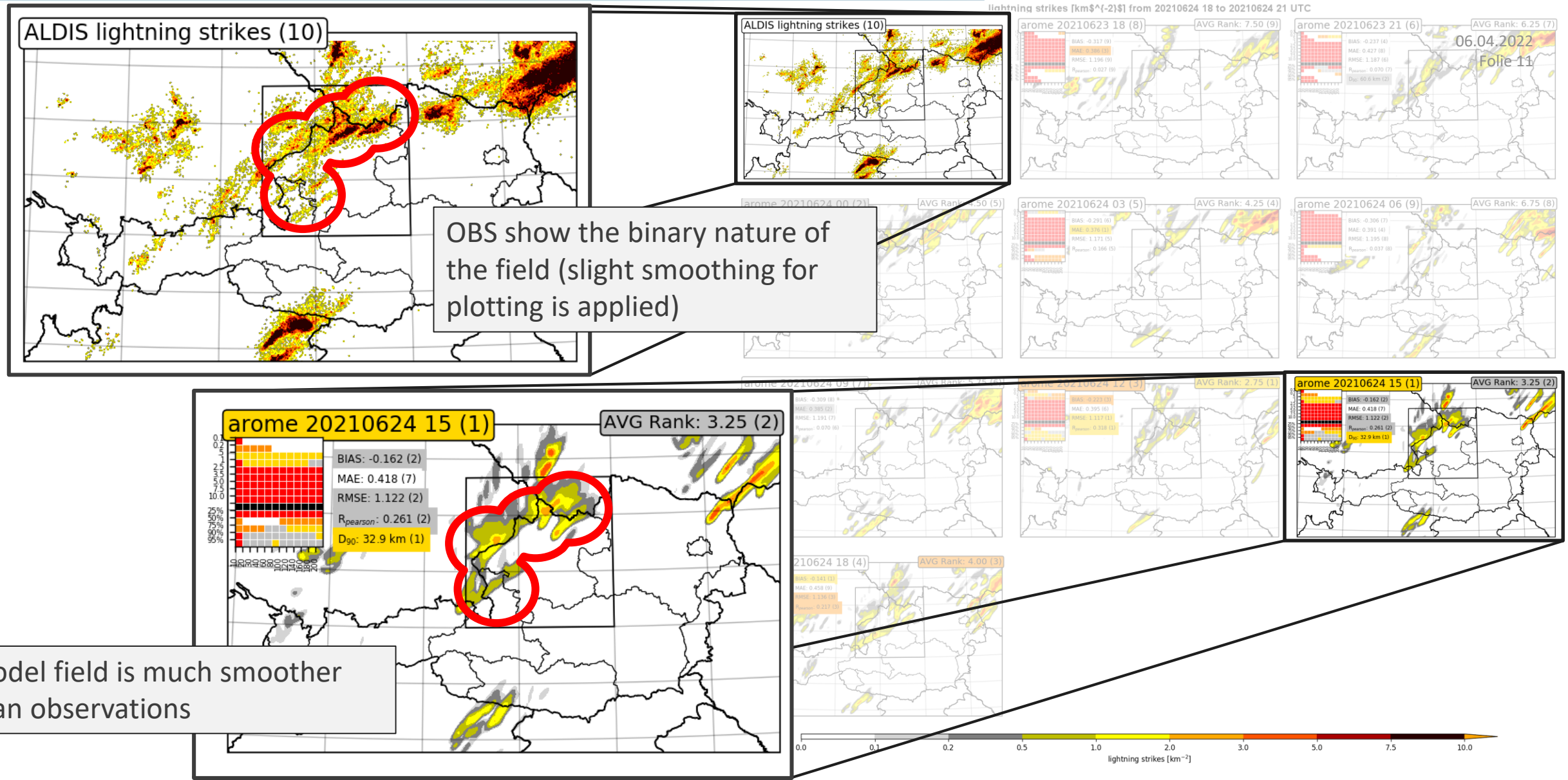
ZAMG
Zentralanstalt für Meteorologie und Geodynamik

Example for lightning verification

- Lightning OBS and model data is presented together
- Scores/Ranking analogous to precipitation data
- Can be used as a supplement to precipitation verification
- The models seem to underestimate the number of lightning strikes for this example (all biases are negative)



Example for lightning verification

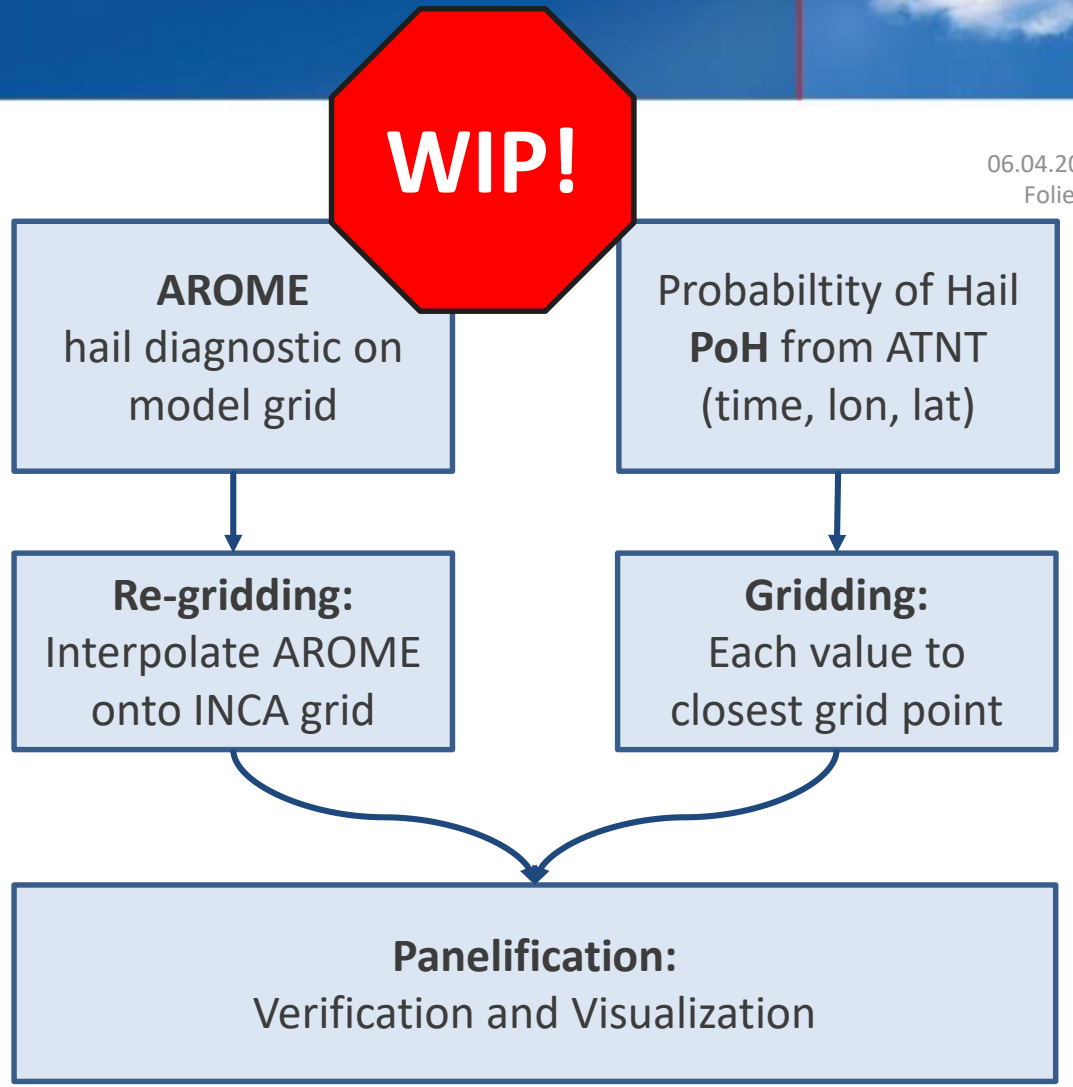


New Variable 2: Hail

- **OBS**: Probability of Hail from ATNT
- **Model**: several options
  - hail diagnostic from the model (SURFDIAGHAIL)
  - Direct hail from ICE-4 or LIMA (currently not used)
  - Calculating PoH from model parameters during runtime

- Advantages:
  - **Highly relevant and impactful** phenomenon -> good choice to optimize for in severe weather forecasting

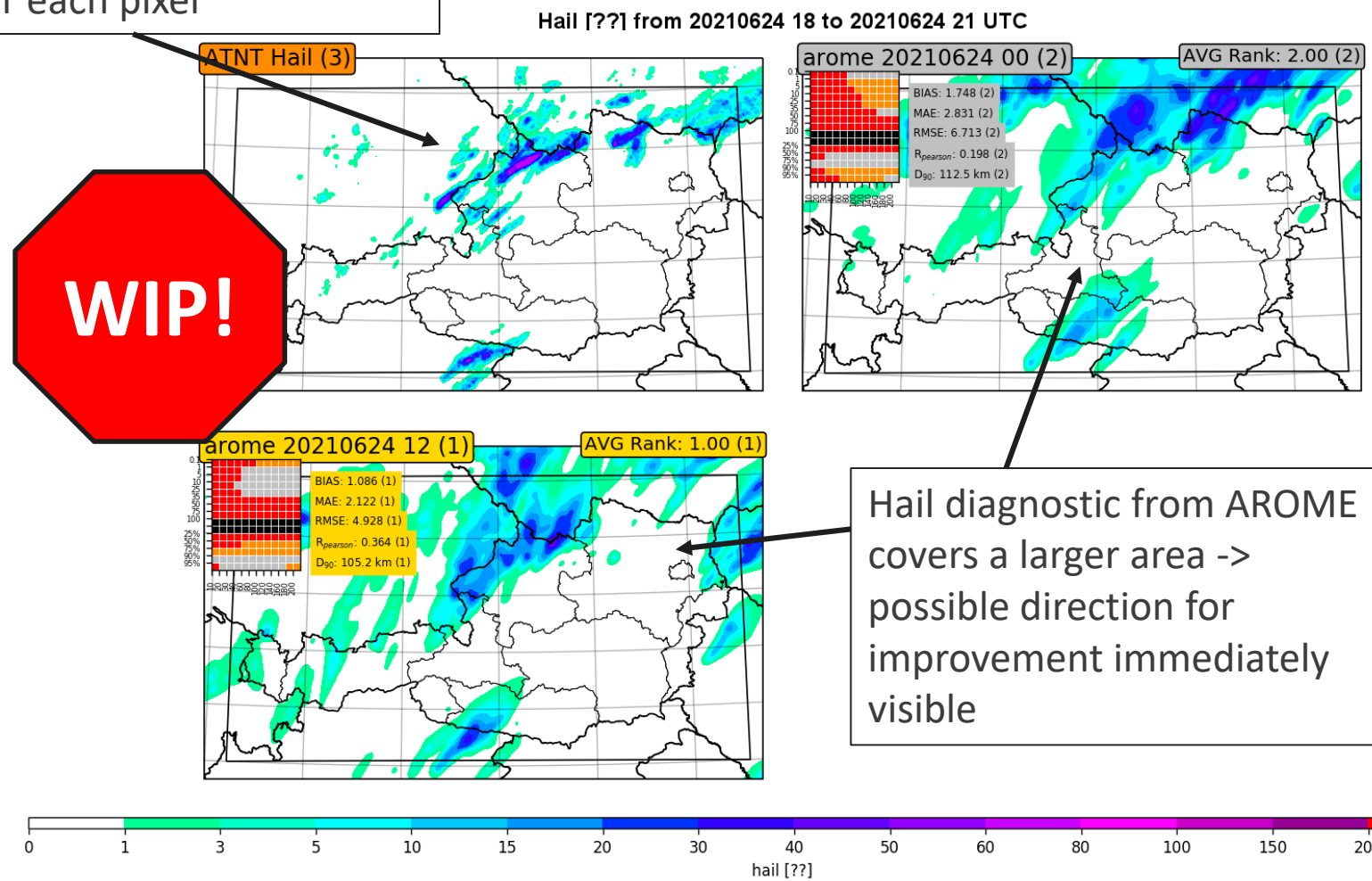
- Caveats:
  - Detection is not straightforward (hail vs graupel vs rain)
  - Currenlty, only the **diagnostic** is available
  - How to compare? PoH vs. kg m^{-2}



ZAMG
Zentralanstalt für Meteorologie und Geodynamik

New Variable 2: Hail (raw values preview)

Max. PoH during the verification period for each pixel



Hail diagnostic from AROME covers a larger area -> possible direction for improvement immediately visible

TODO:

- Use warning thresholds for light, moderate, severe hail as used in ZAMG hail warnings
- Use equivalent thresholds for PoH
- Compare the resulting fields to prevent comparing apples to oranges

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

Overview

What is the challenge?

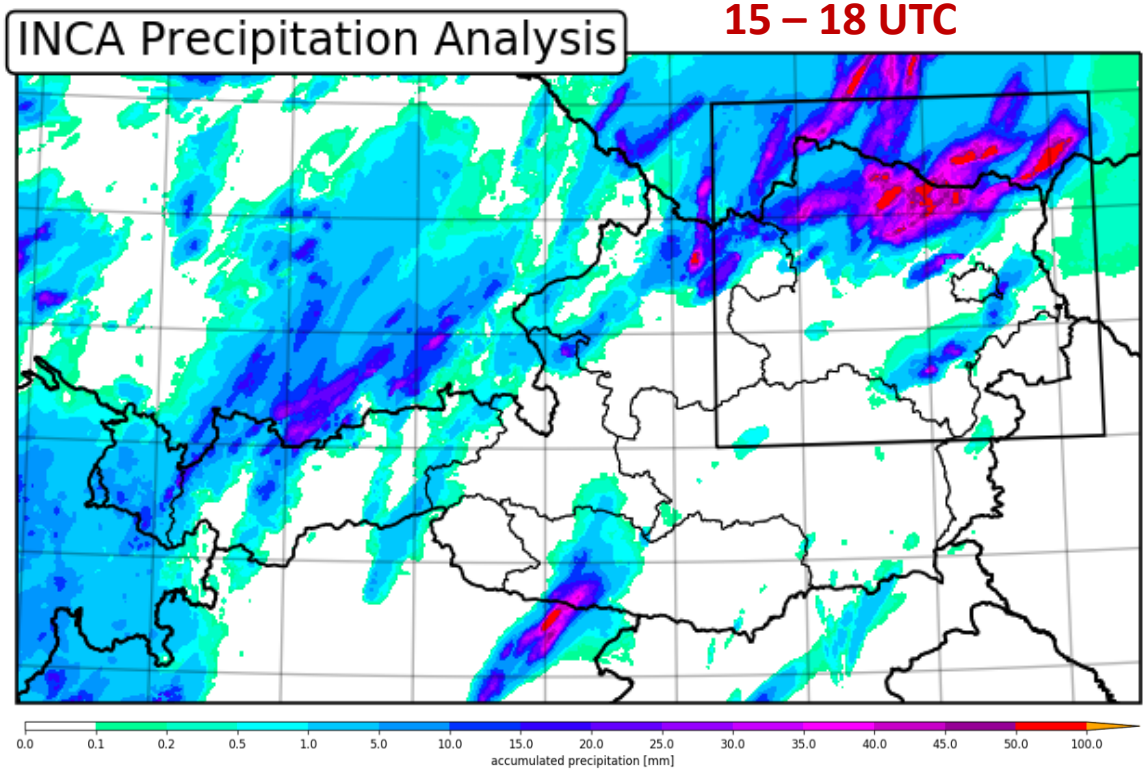
Upcoming Parameters

A Look at the Severe Storm of 24 June 2021

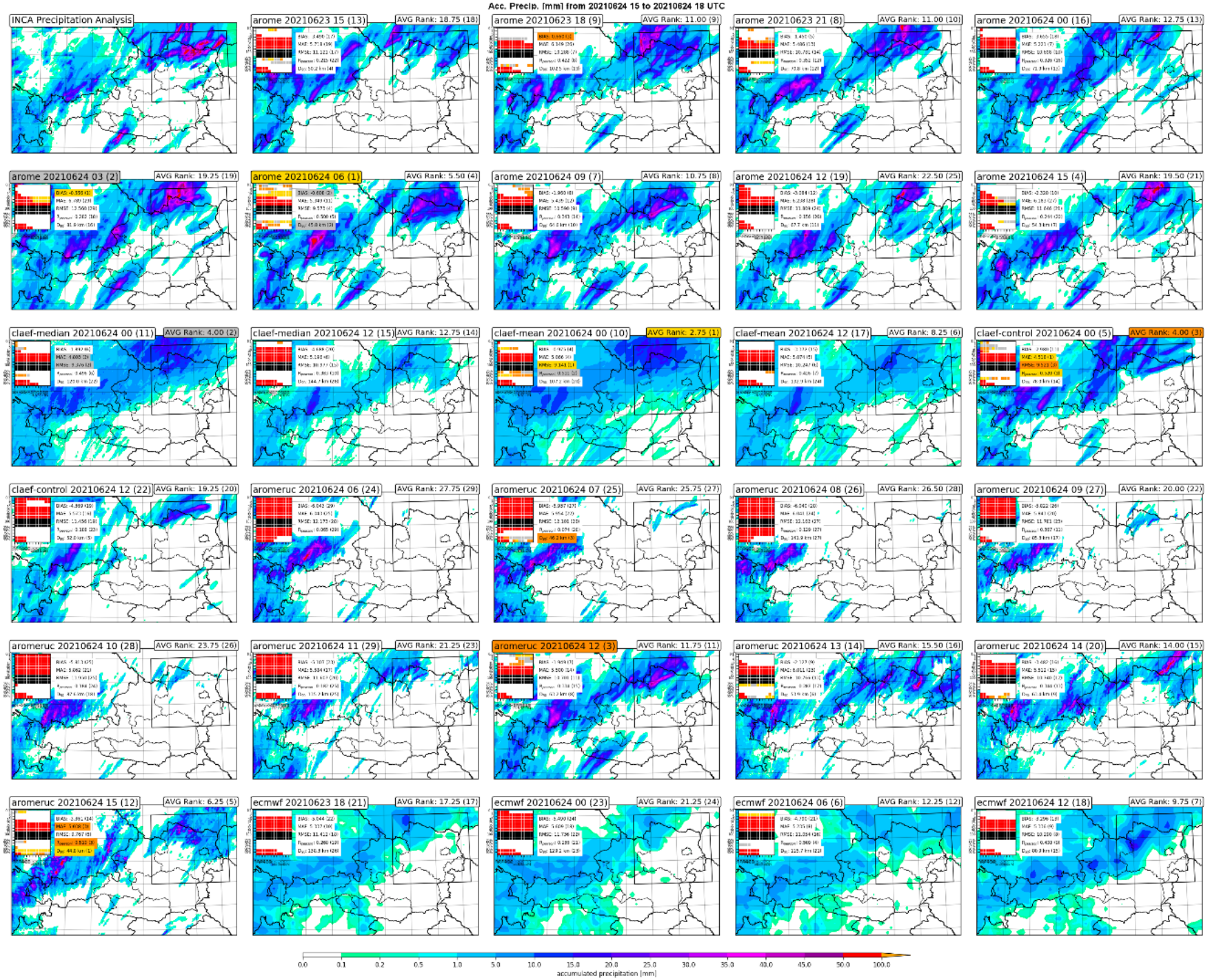
24 June 2021: Hail & Tornado along the Austrian-Czech Boarder

Hail in Lower Austria
(APA/Feuerwehr)

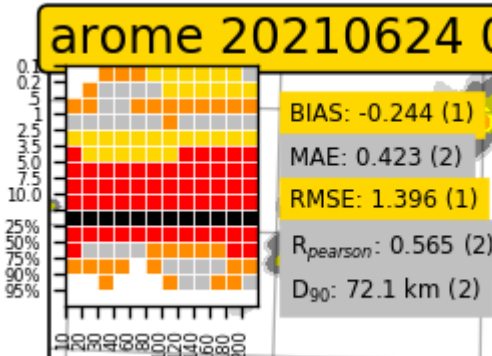
Tornado damage in Hrušky, CZ (imago images/CTK Photo)

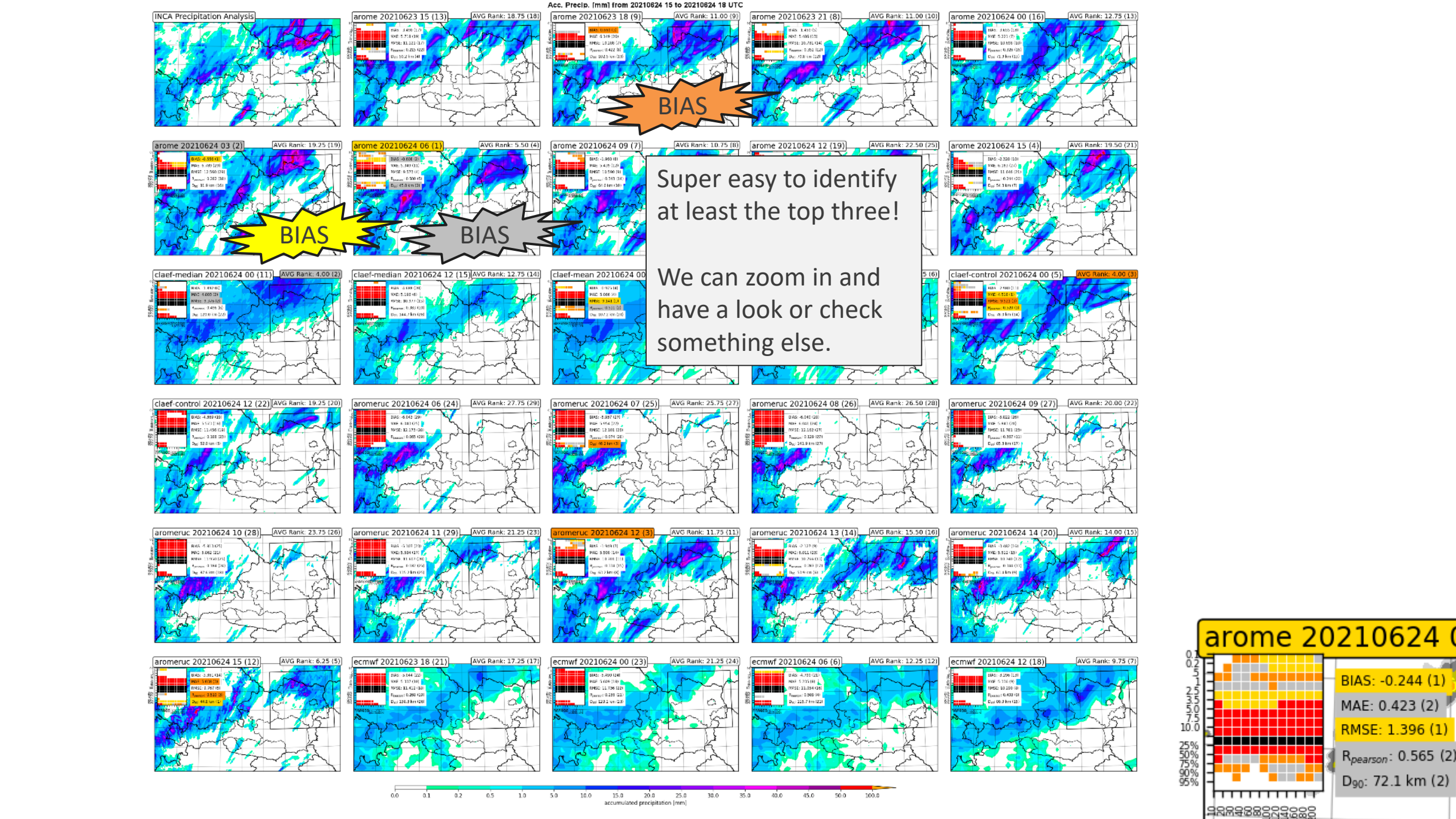


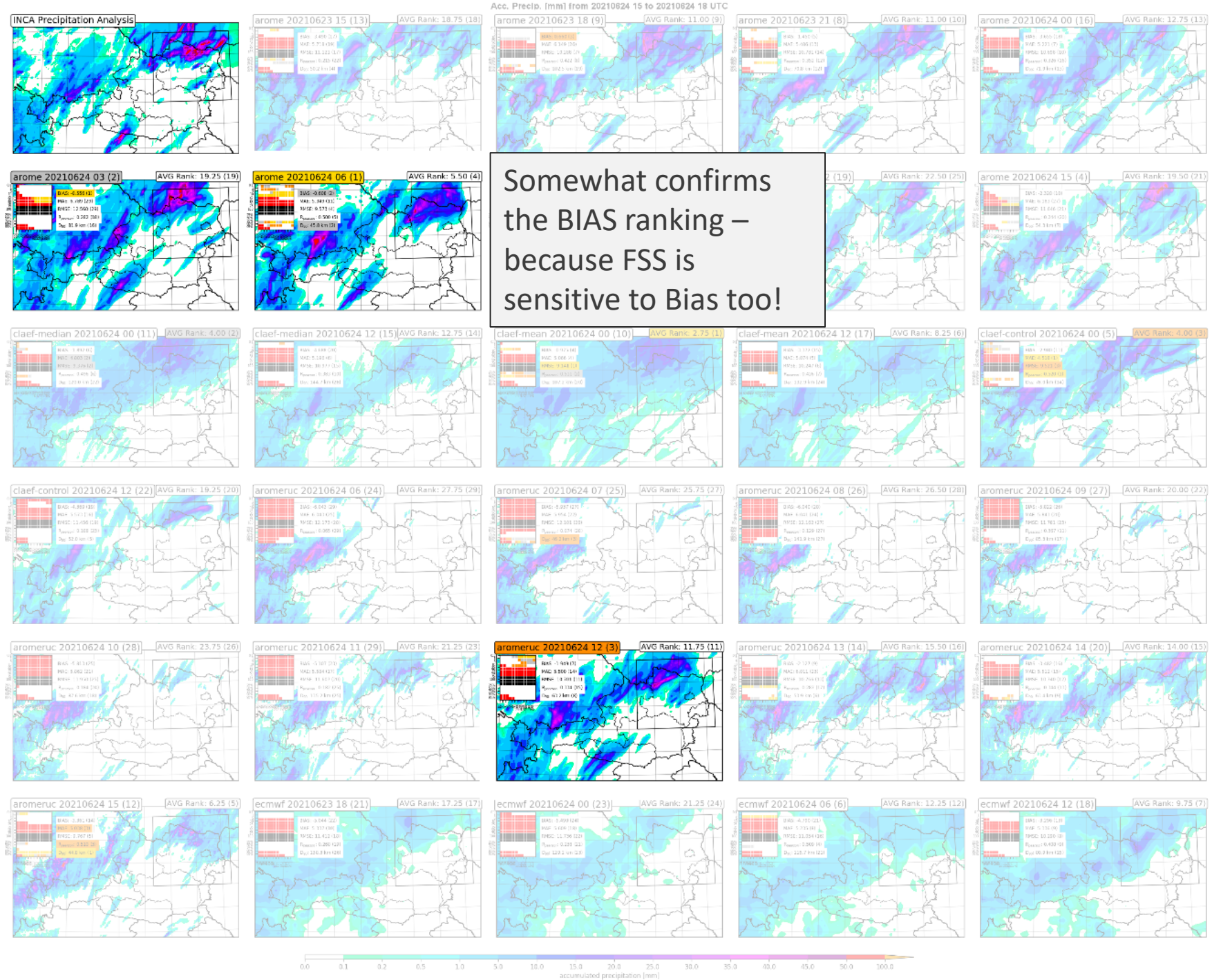
ZAMG
Zentralanstalt für Meteorologie und Geodynamik



Quick reminder below, which score is where?

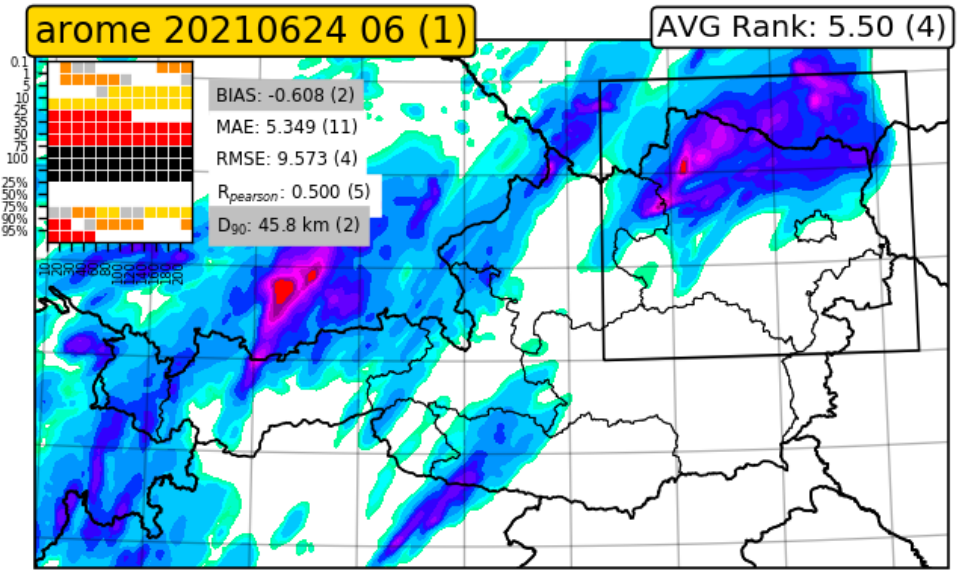


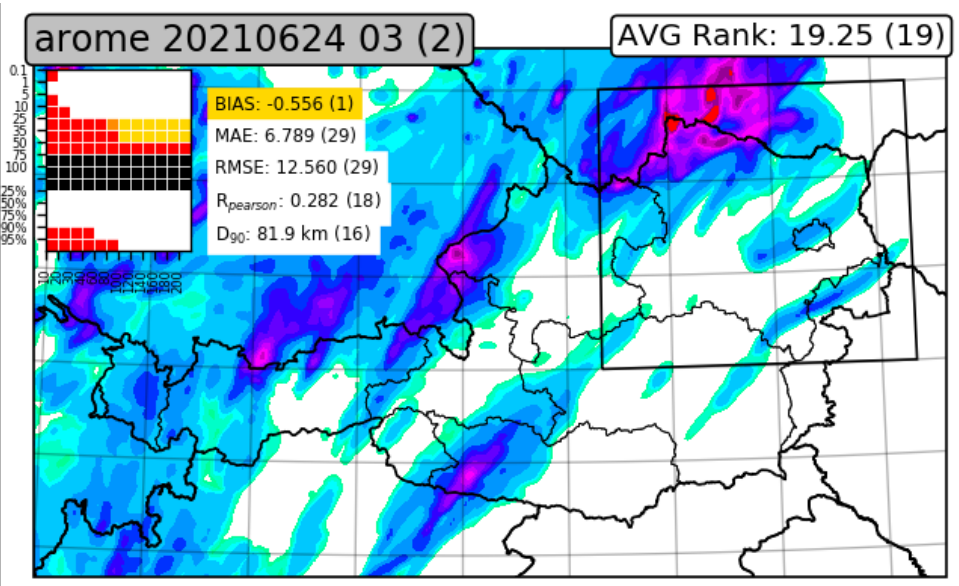

Acc. Precip. [mm] from 20210624 15 to 20210624 18 UTC
INCA Precipitation Analysis
arome 20210623 15 (13)
AVG Rank: 18.75 (18)
arome 20210623 18 (9)
AVG Rank: 11.00 (9)
arome 20210623 21 (8)
AVG Rank: 11.00 (10)
arome 20210624 00 (16)
AVG Rank: 12.75 (13)
BIAS
arome 20210624 03 (2)
AVG Rank: 19.25 (19)
arome 20210624 06 (1)
AVG Rank: 5.50 (4)
arome 20210624 09 (7)
AVG Rank: 10.75 (8)
arome 20210624 12 (19)
AVG Rank: 22.50 (25)
arome 20210624 15 (4)
AVG Rank: 19.50 (21)
BIAS
BIAS
Super easy to identify at least the top three!
We can zoom in and have a look or check something else.
claef-median 20210624 00 (11)
AVG Rank: 4.00 (2)
claef-median 20210624 12 (15)
AVG Rank: 12.75 (14)
claef-mean 20210624 00
claef-control 20210624 00 (5)
AVG Rank: 4.00 (3)
claef-control 20210624 12 (22)
AVG Rank: 19.25 (20)
aromeruc 20210624 06 (24)
AVG Rank: 27.75 (29)
aromeruc 20210624 07 (25)
AVG Rank: 25.75 (27)
aromeruc 20210624 08 (26)
AVG Rank: 26.50 (28)
aromeruc 20210624 09 (27)
AVG Rank: 20.00 (22)
aromeruc 20210624 10 (28)
AVG Rank: 23.75 (26)
aromeruc 20210624 11 (29)
AVG Rank: 21.25 (23)
aromeruc 20210624 12 (3)
AVG Rank: 11.75 (11)
aromeruc 20210624 13 (14)
AVG Rank: 15.50 (16)
aromeruc 20210624 14 (20)
AVG Rank: 14.00 (15)
aromeruc 20210624 15 (12)
AVG Rank: 6.25 (5)
ecmwf 20210623 18 (21)
AVG Rank: 17.25 (17)
ecmwf 20210624 00 (23)
AVG Rank: 21.25 (24)
ecmwf 20210624 06 (6)
AVG Rank: 12.25 (12)
ecmwf 20210624 12 (18)
AVG Rank: 9.75 (7)
0.0 0.1 0.2 0.5 1.0 5.0 10.0 15.0 20.0 25.0 30.0 35.0 40.0 45.0 50.0 100.0
accumulated precipitation [mm]
arome 20210624
BIAS: -0.244 (1)
MAE: 0.423 (2)
RMSE: 1.396 (1)
$R_{pearson}$: 0.565 (2)
D_{90}: 72.1 km (2)
0.1 0.2 5 1 2.5 3.5 5.0 7.5 10.0
25% 50% 75% 90% 95%

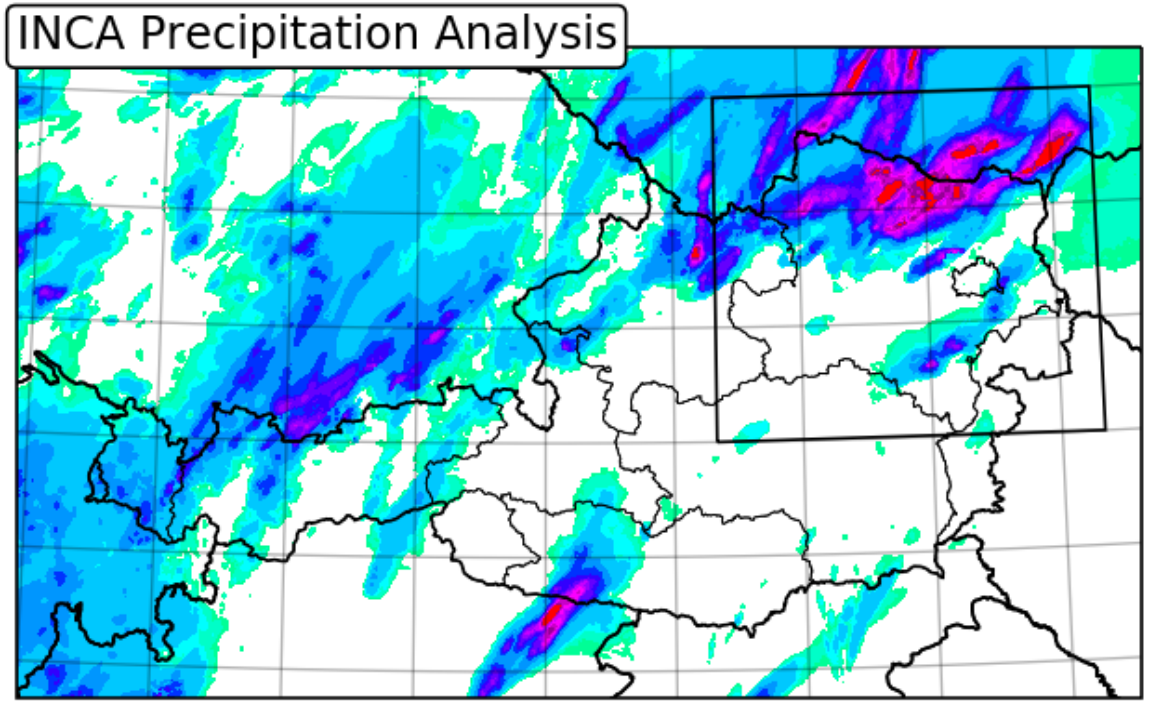


Somewhat confirms the BIAS ranking – because FSS is sensitive to Bias too!

A closer look at the winners



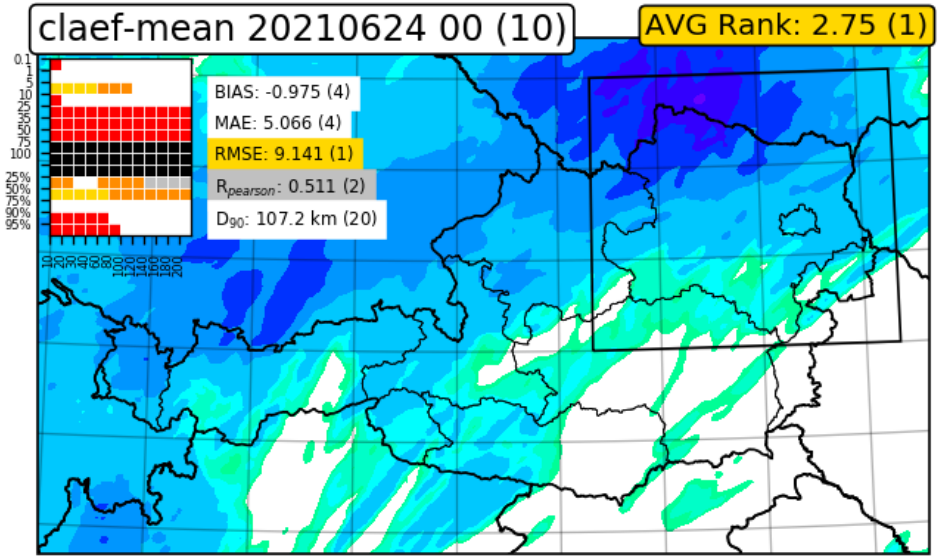


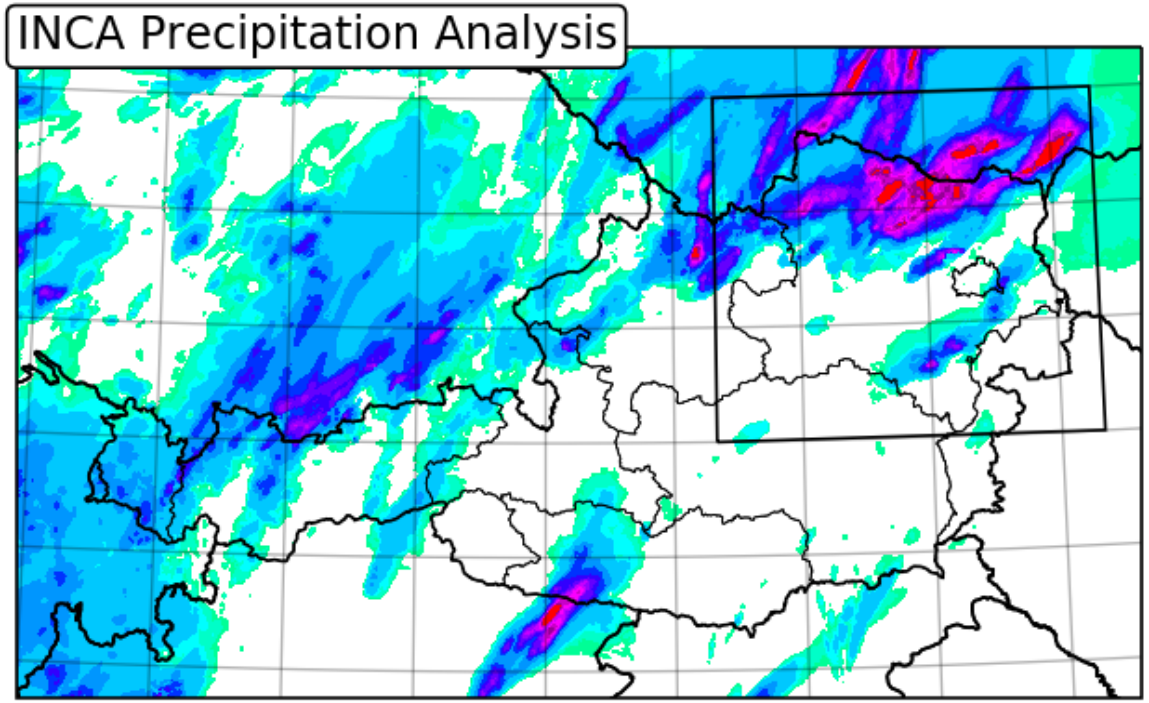


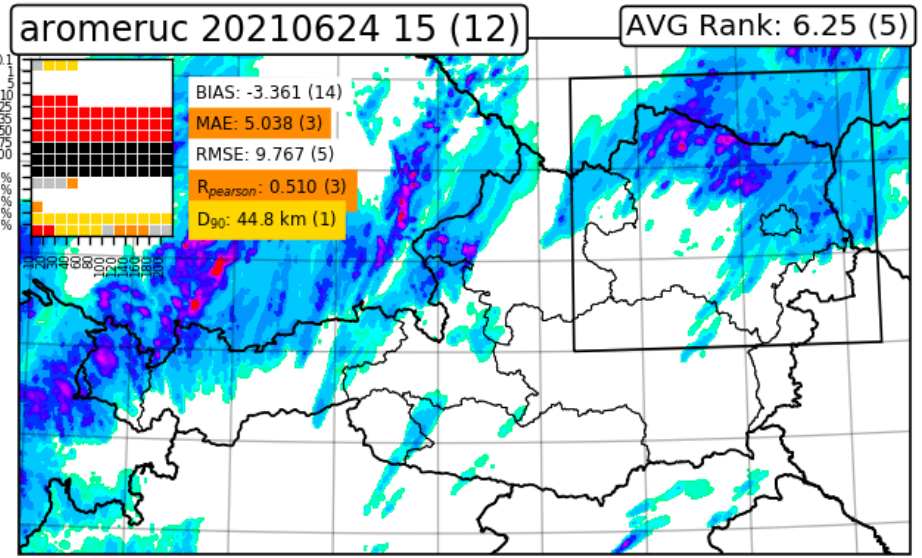
- FSS Rank Score identifes two simulations with good overall precipitation distribution and reasonable bias
- Rank 1 also scores highly in RMSE, Correlation and D90

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

What else can we learn? Two quick examples





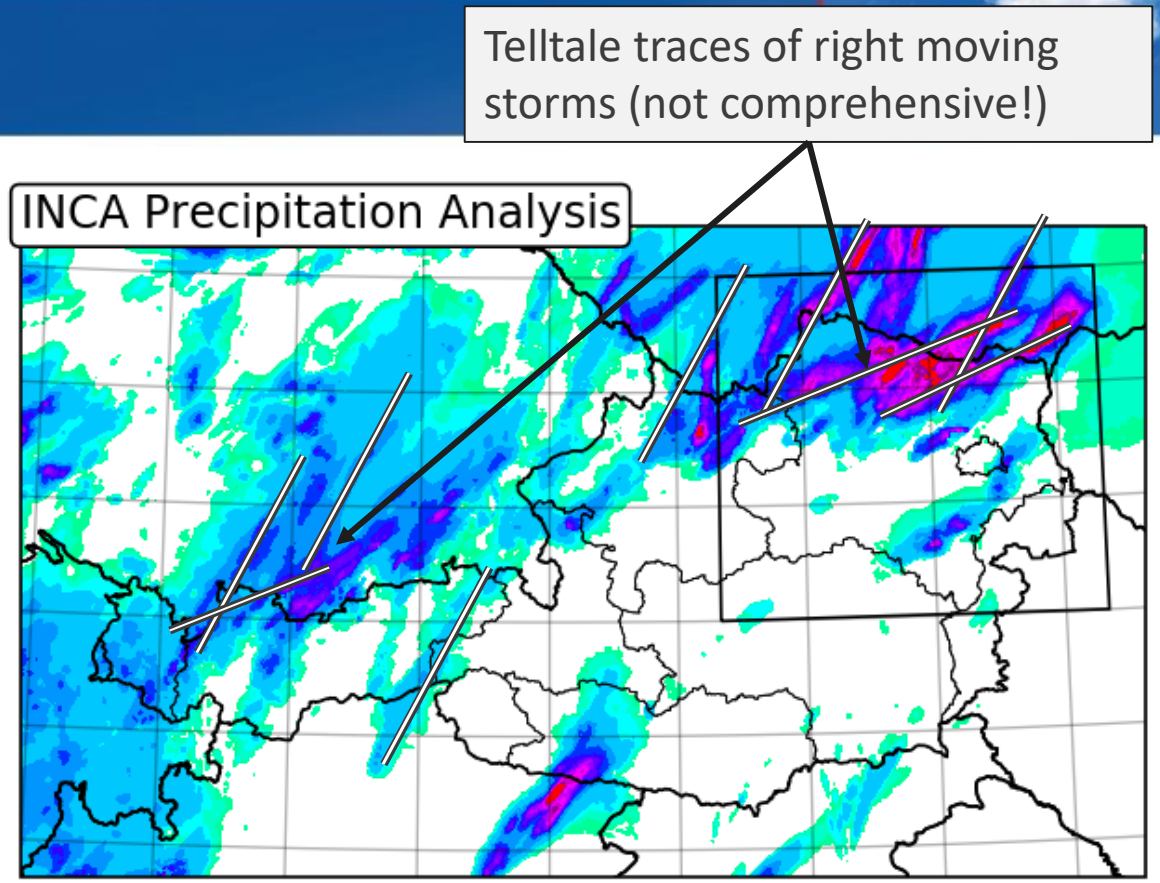


- Average of classic scores tends to favor diffuse precipitations fields, especially RMSE and MAE often are low for global models
- D90 is low (good) for where extreme values are within close proximity in OBS and model

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

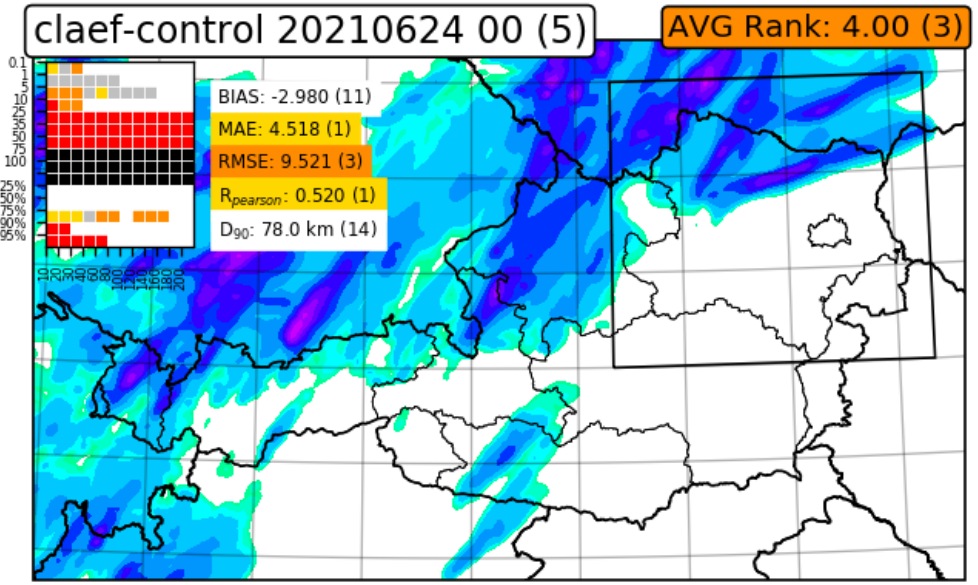
But wait, there‘s more!

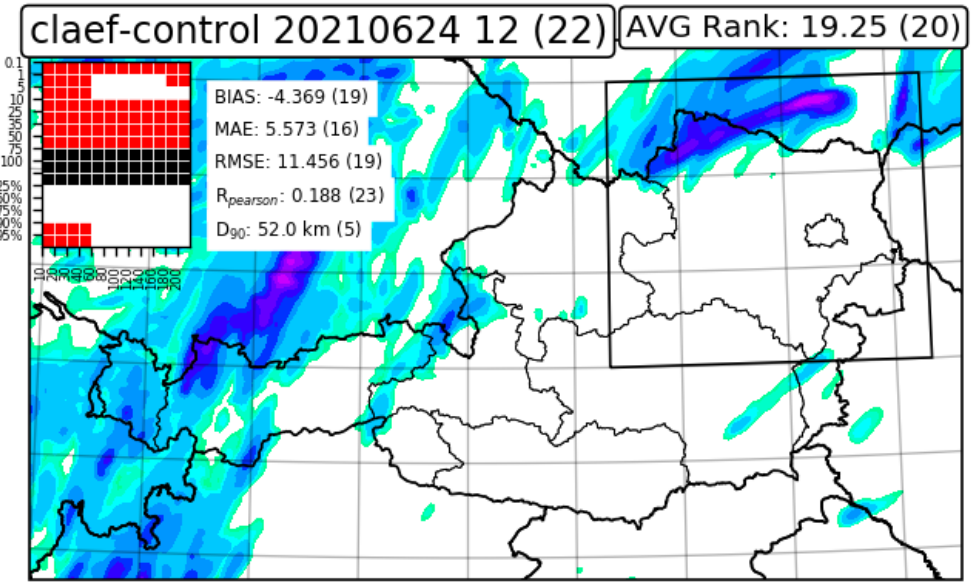
- Visualization is essential for this process!
- Even a quick examination allows to find several traces of moving cells in the observations. This is by no means perfect, but it‘s a good starting point!
- We can look for such traces in model fields and try to find supercells in the models

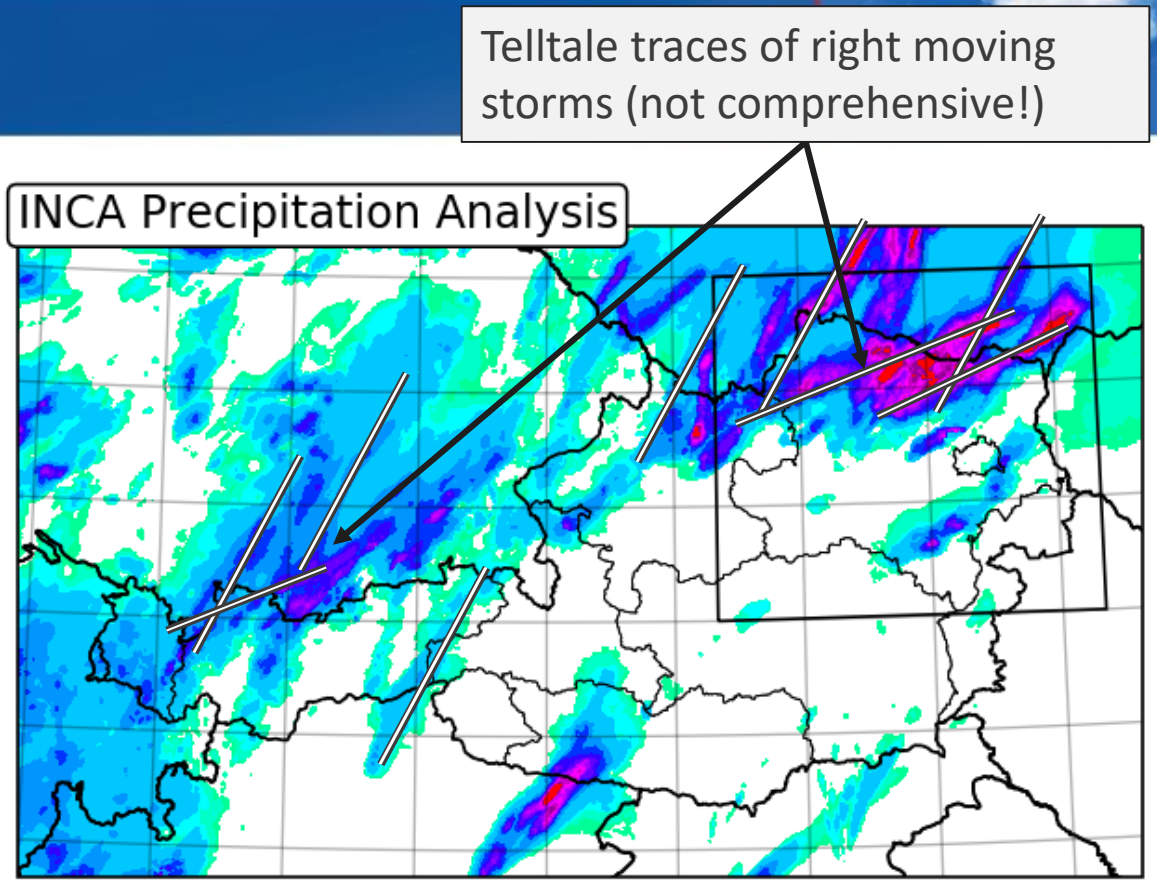


ZAMG
Zentralanstalt für Meteorologie und Geodynamik

Interesting candidates for a case study?







- For example: **visual examination** lets us easily identify two simulations with **pronounced supercell signatures**
- Could serve as a **starting point for analyzing** storm cell dynamics
- Entire **plotting and analyzing** of this example is doable in **less than 20 minutes by a single person**

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

Closing Remarks and Outlook

- Panelification has become a valuable tool to gain a quick overview on model performance after warning cases and severe storms

- Lightning and hail will soon be fully implemented as verification parameters

Outlook

- Continue to optimize Panelification based on the input of experts (new scores, better ranking, other options?)

- Possibly implement a **similar visualization with HAARP**?

- Continue and expand this human-centered approach to verification?

Appendix I: Hail Calculations for OBS and Model

Hail kinetic energy flux → Severe hail index **SHI** → Maximum expected hail size **MEHS**

hail kinetic energy ($\dot{E}$) (Waldvogel et al. 1978a; Waldvogel et al. 1978b; Federer et al. 1986) by

$$\dot{E} = 5 \times 10^{-6} \times 10^{0.084Z} W(Z), \qquad (1)$$

where

$$W(Z) = \begin{cases} 0 & \text{for } Z \le Z_L \\ \dfrac{Z - Z_L}{Z_U - Z_L} & \text{for } Z_L < Z < Z_U \\ 1 & \text{for } Z \ge Z_U. \end{cases}$$

Here Z is in dBZ, $\dot{E}$ in Joules per square meter per second, and the weighting function $W(Z)$ can be used to define a transition zone between rain and hail reflectivities. The default values for this algorithm have initially been set to $Z_L = 40$ dBZ and $Z_U = 50$ dBZ (but are adaptable).[1]

$$W_T(H) = \begin{cases} 0 & \text{for } H \le H_0 \\ \dfrac{H - H_0}{H_{m20} - H_0} & \text{for } H_0 < H < H_{m20} \\ 1 & \text{for } H \ge H_{m20}, \end{cases} \qquad (2)$$

where H is the height above radar level (ARL), H_0 is the height ARL of the environmental melting level, and H_{m20} is the height ARL of the −20°C environmental temperature. Both H_0 and H_{m20} can be determined from a nearby sounding or from other sources of upper-air data (e.g., numerical model output).

$$\mathrm{SHI} = 0.1 \int_{H_0}^{H_T} W_T(H)\dot{E}\, dH, \qquad (3)$$

$$\mathrm{MEHS} = 2.54(\mathrm{SHI})^{0.5}$$

Probability of hail **PoH**

$$POH = 100 * \left(\frac{\tanh\left(\frac{MEHS - 25.0}{15.9155}\right)}{1.8} + 0.5\right)$$

SURFDIAGHAIL – Hail from AROME

Maximum of the vertically integrated graupel content between output time steps

Witt, A., Eilts, M. D., Stumpf, G. J., Johnson, J. T., Mitchell, E. D. W., & Thomas, K. W. (1998). An Enhanced Hail Detection Algorithm for the WSR-88D, *Weather and Forecasting*, *13*(2), 286-303. Retrieved Apr 1, 2022, from https://journals.ametsoc.org/view/journals/wefo/13/2/1520-0434_1998_013_0286_aehdaf_2_0_co_2.xml

ZAMG Zentralanstalt für Meteorologie und Geodynamik

Appendix II: Lightning Diagnostic in AROME

Lightning diagnostic in AROME

005), we propose that one useful estimate of the total flash rate may be based on the resolved upward flux wq_g of large precipitating ice (i.e., graupel) in the mixed-phase region at −15°C. We designate this first type of threat estimate by the symbol F_1. For this threat we thus assume

$$F_1 = f[(wq_g)_m], \quad (1)$$

where w is the vertical velocity, q_g is the graupel mixing ratio, and the subscript m attached to the flux implies evaluation at the −15°C level in the mixed-phase region.

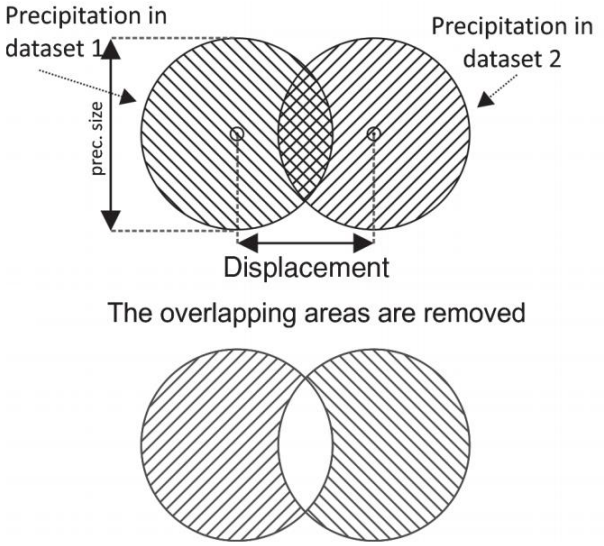
For AROME Aut, the value was adjusted for several severe storm events to obtain a good estimate of the total amount of lightning strikes

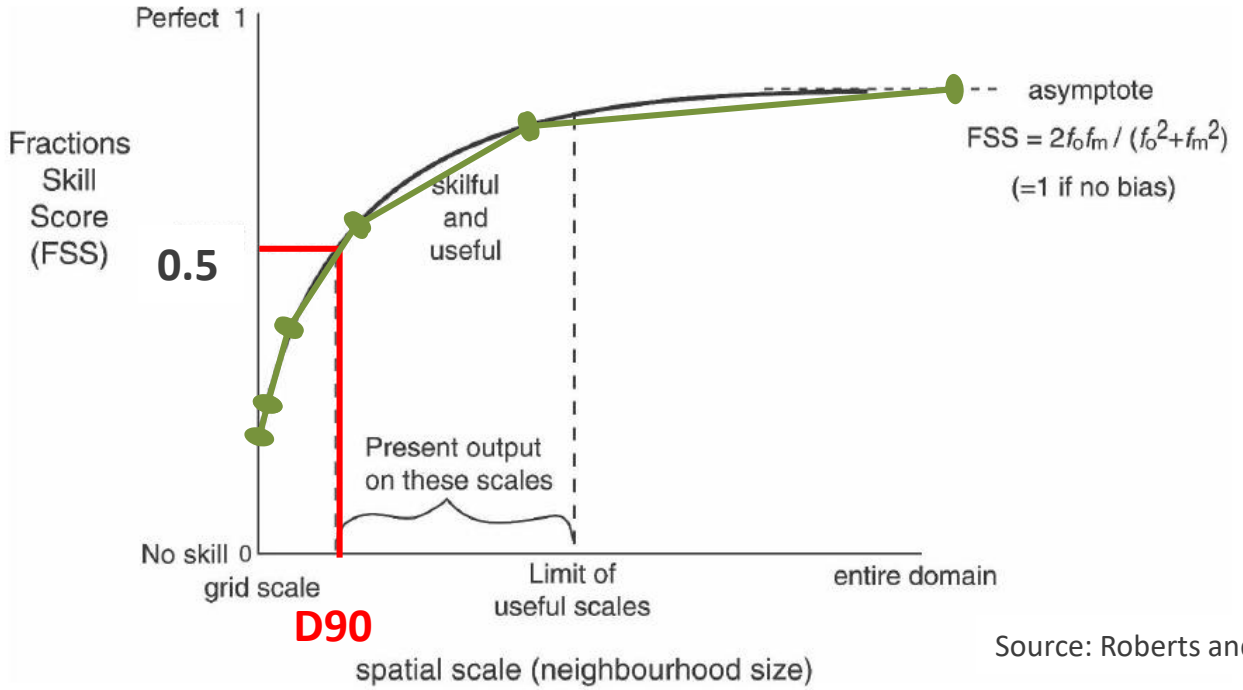
McCaul , E. W., Jr., Goodman, S. J., LaCasse, K. M., & Cecil, D. J. (2009). Forecasting Lightning Threat Using Cloud-Resolving Model Simulations, *Weather and Forecasting*, *24*(3), 709-729. Retrieved Apr 1, 2022, from https://journals.ametsoc.org/view/journals/wefo/24/3/2008waf2222152_1.xml

ZAMG Zentralanstalt für Meteorologie und Geodynamik

Appendix III: D90 - Displacement of the 90th precipitation percentile

- Use 90th Percentile -> removes bias
- D90 is defined as the window size at which the FSS exceeds 0.5, the threshold for a skillful and useful forecast





Approximation:

1. Remove Overlap
2. Calculate FSS for 1, 2, 4, 8, … 2k windows
3. Stop when FSS > 0.5
4. Linearly interpolate to 0.5

Source: Roberts and Lean (2007), **Skok and Roberts (2018)**

ZAMG Zentralanstalt für Meteorologie und Geodynamik