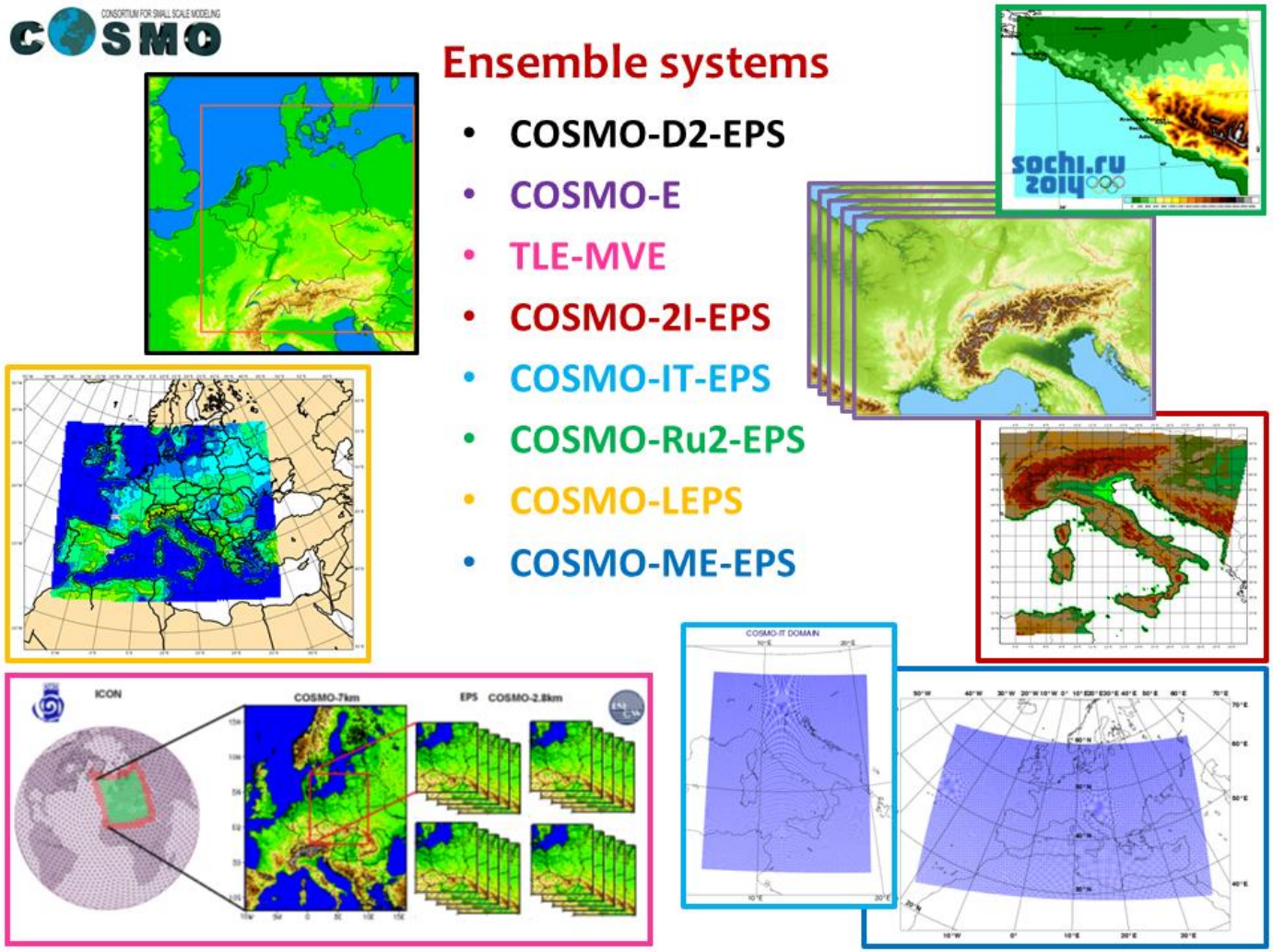
COSMO ensembles: Overview and lessons learned

Chiara Marsigli
Deutscher Wetterdienst

The COSMO ensembles



http://www.cosmo-model.org/content/tasks/workGroups/wg7

The COSMO ensembles

- the COSMO members develop and maintain several ensemble systems at the convection-permitting scale:
  - COSMO-D2-EPS, by DWD, operational, 2.2 km
  - COSMO-E, by MCH, operational, 2.2 km
  - TLE-MVE, by IMGW, operational, 2.8 km
  - COSMO-2I-EPS, by Arpae, pre-operational, 2.2 km
  - COSMO-IT-EPS, by COMET, pre-operational, 2.2 km
  - COSMO-Ru2-EPS, by RHM, for research, 2.8 km

- COMET operates an ensemble at 7 km, COSMO-ME-EPS

- COSMO-LEPS is the Consortium ensemble, running since 2002, 7 km

The problem of the new coffee machine
(or: Pros and cons of ensemble forecasting)

- How to use a new (highly technical!) coffee machine?
  - One run only, with a default set-up of the machine parameters, does not give a satisfactory output
  - Make many runs, by varying the machine parameters (% of milk, strength of the coffee, total amount, …): too much output!

- How to benefit from 20 different coffees?
  - Ensemble mean
  - Clusters
  - Extremes
  - Select the best
  - Rank the frequencies

Outline

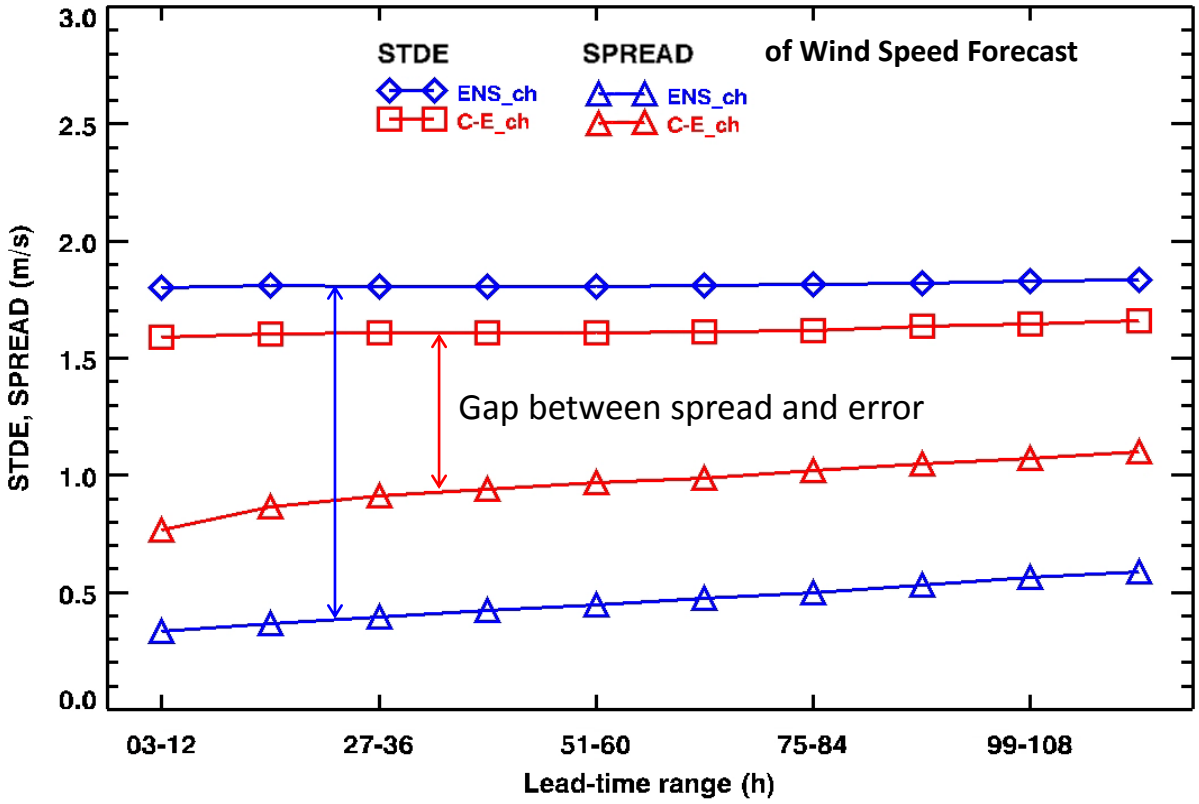
- Initial conditions
- Spread/skill relation
- Model bias
- Ensemble products
- Verification adapted to the model skill

Initial conditions

- Initial conditions for the COSMO ensembles provided by KENDA analyses at:
  - DWD
  - MeteoSwiss
  - Arpae
  - COMET
- Perturbed Initial Conditions, thanks to LETKF Data Assimilation
- Is this the “optimal” choice for ensemble forecasting?
- Data assimilation requirements are not the same as ensemble forecast requirements
- The ensemble which is suitable for data assimilation may not the same which is suitable for weather forecast (e.g. spread) -> two problems with **almost** the same solution
- Discuss with Data Assimilation group

Spread/skill relation

COSMO-E vs IFS-ENS over Switzerland for summer (JJA) 2018



P. Kaufmann, A. Walser - MCH

The problem of the model bias

TIGGE-LAM

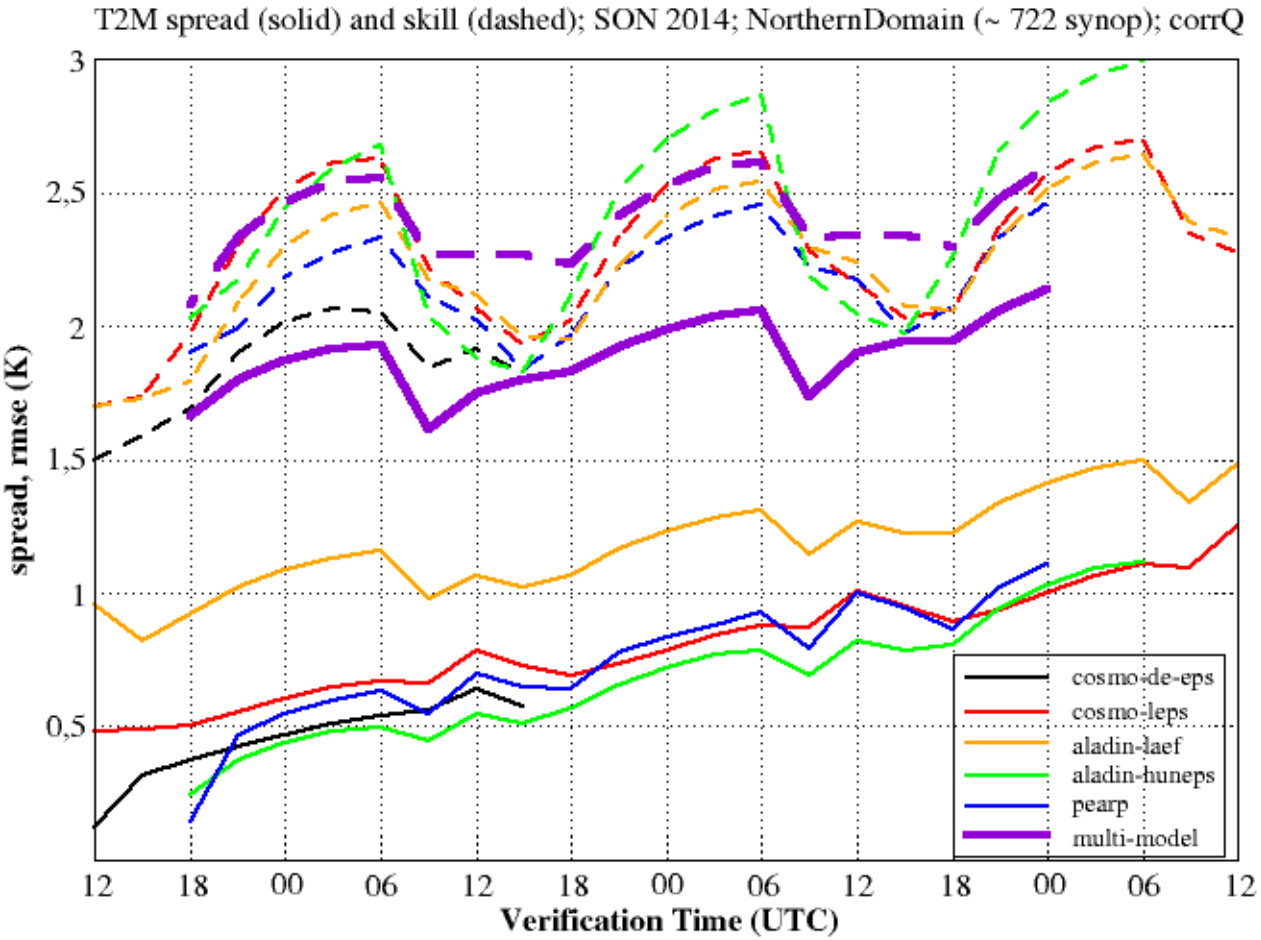
COSMO-DE-EPS (20 m, 2.8 km)

COSMO-LEPS (16 m, 7 km)

ALADIN-LAEF (17 m, 15 km)

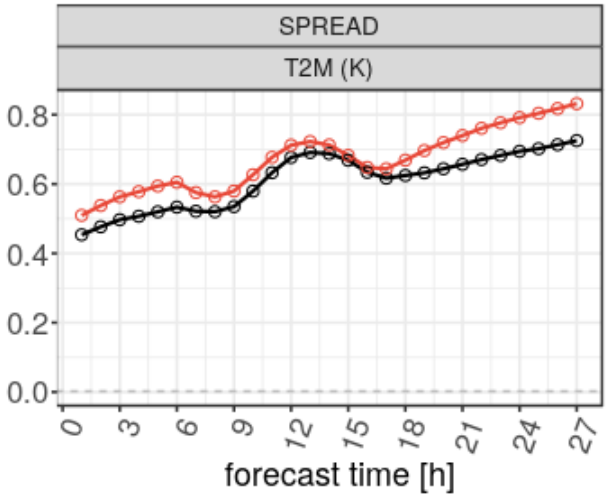
ALADIN-HUN (11 m, 11 km)

PEARP (35 m, 25km)



A. Montani, Arpae-SIMC

How to increase the ensemble spread?



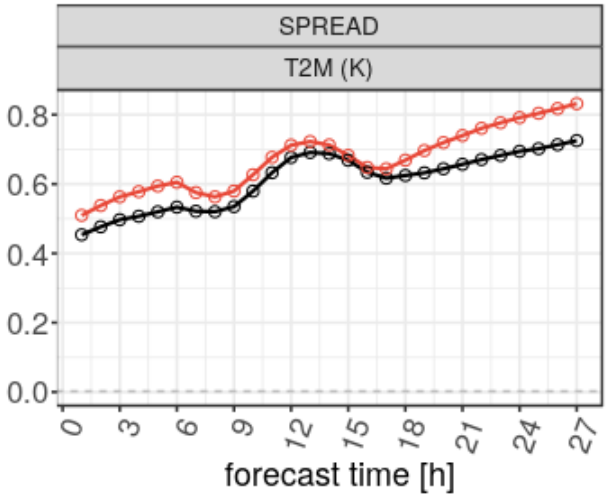
control

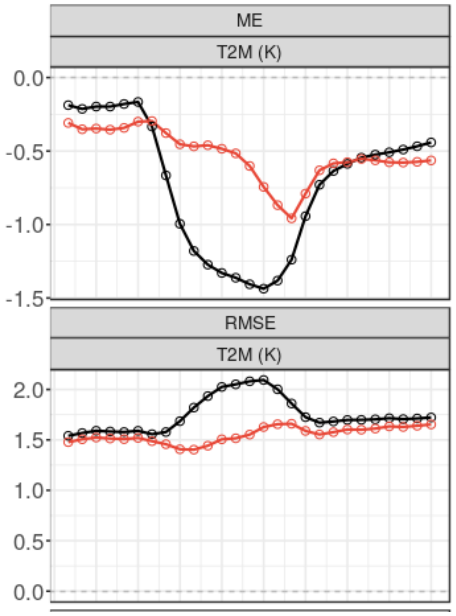
experiment

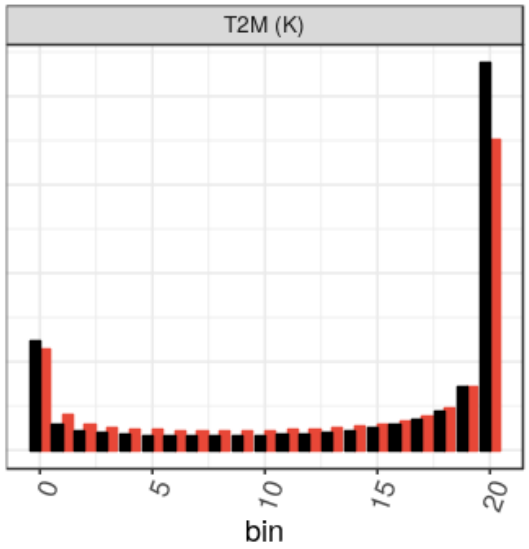
- SPPT?
- Parameter perturbation?
- Stochastic physics?
- Multi-physics?

How to increase the ensemble spread?

- Effect of reduction of the systematic model error on ensemble spread







control

experiment

J.-P. Schulz: Improved land surface processes

Design of ensemble products

- It is better a product which is relevant for the user or a product which the ensemble can provide accurately?

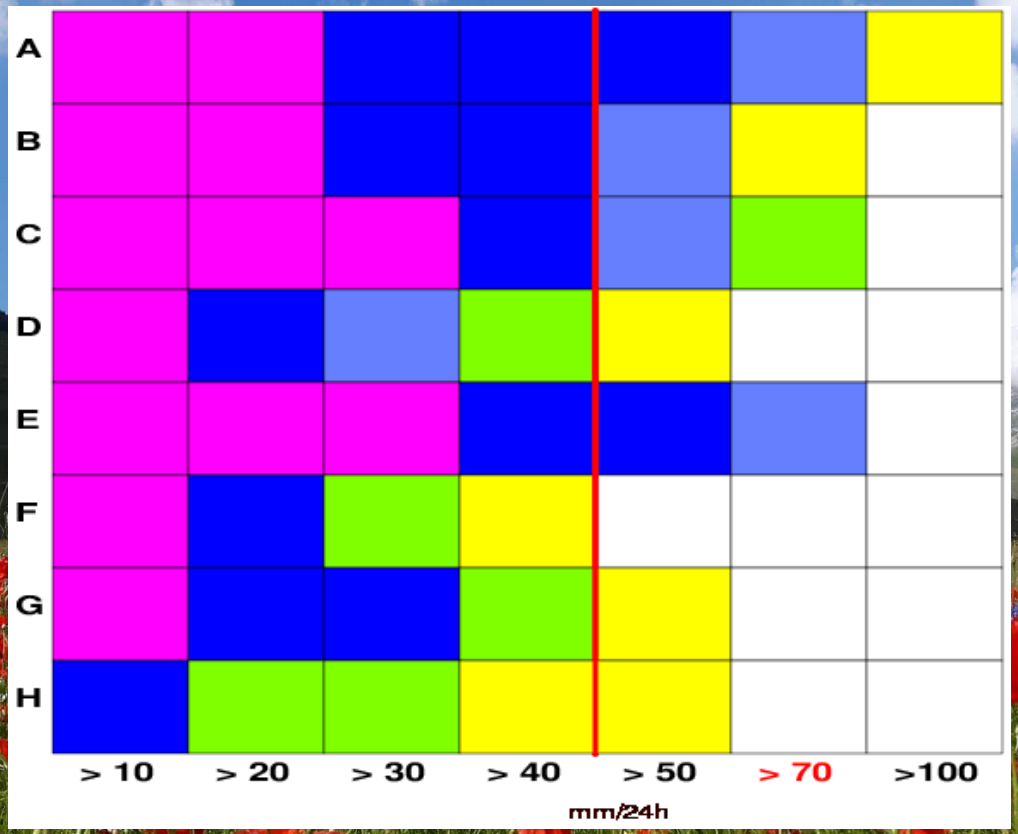
Design of ensemble products

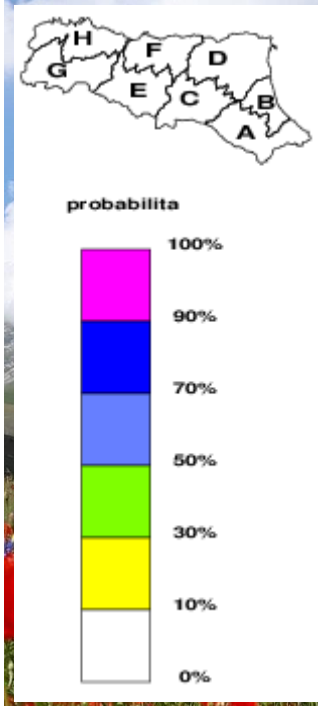
- Use of spatial verification methods for selecting the aggregation scale: how?
  - Select **a level of agreement** between forecast and observations (based on score) and ask at what neighbourhood size this agreement is obtained

  or:

  - Select a neighbourhood size based on the **forecaster need** (catchment, warning area) and assess which forecast is more reliable at that scale
- Products in terms of upscaled quantity
  - Loosing the high-resolution for gaining reliability and realism

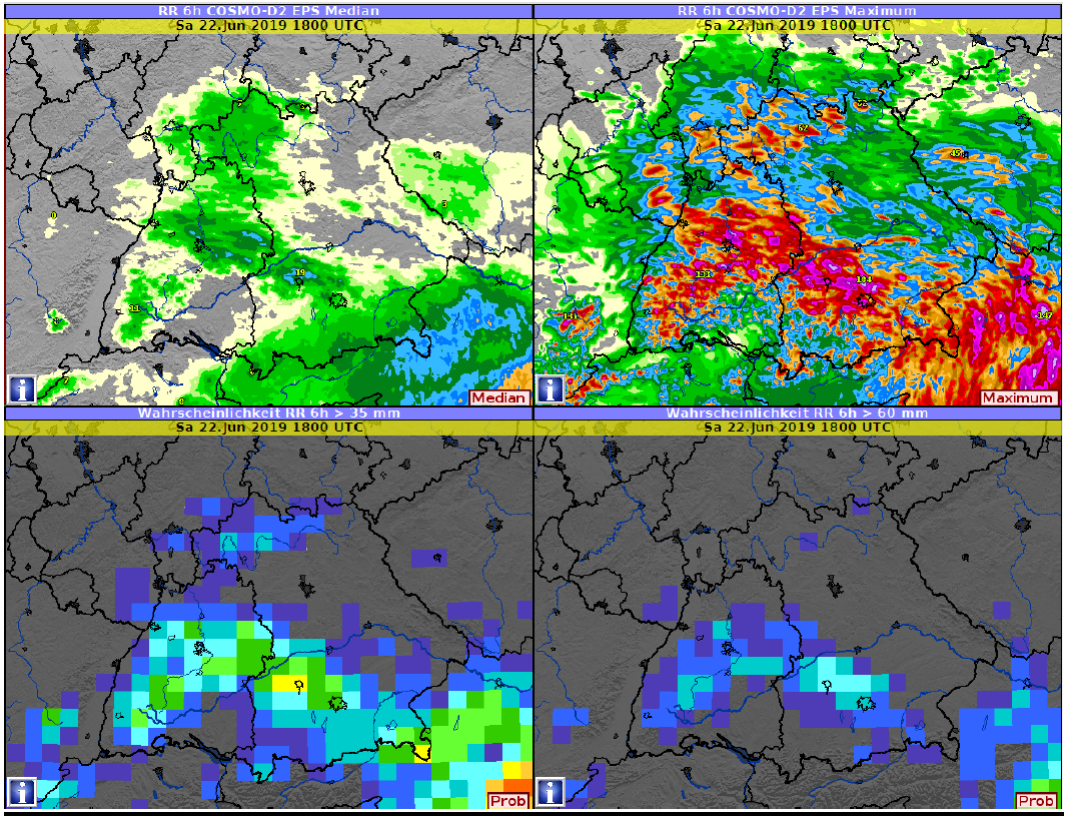
COSMO-LEPS – chessboard

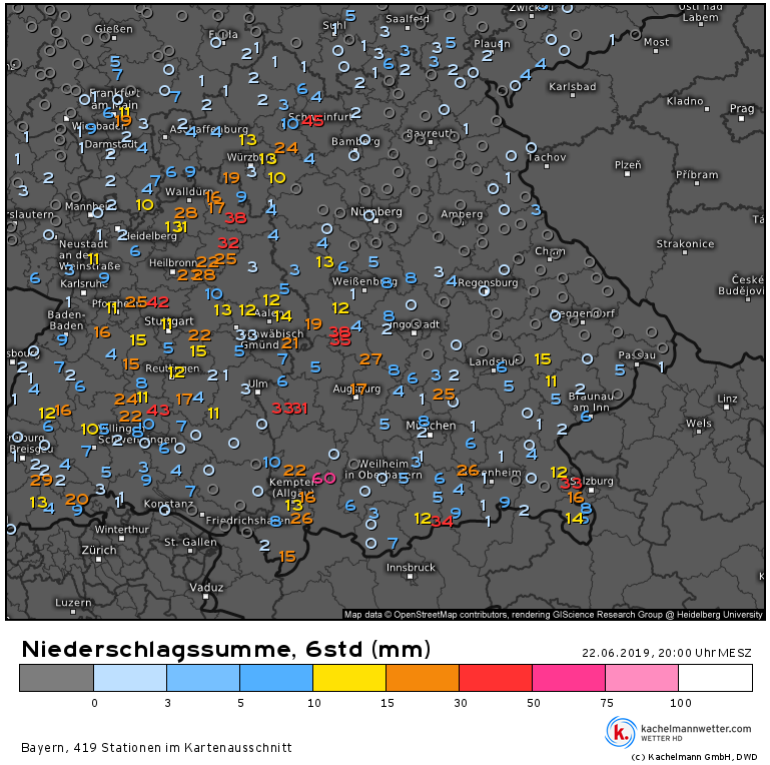




G. Pincini, Arpae

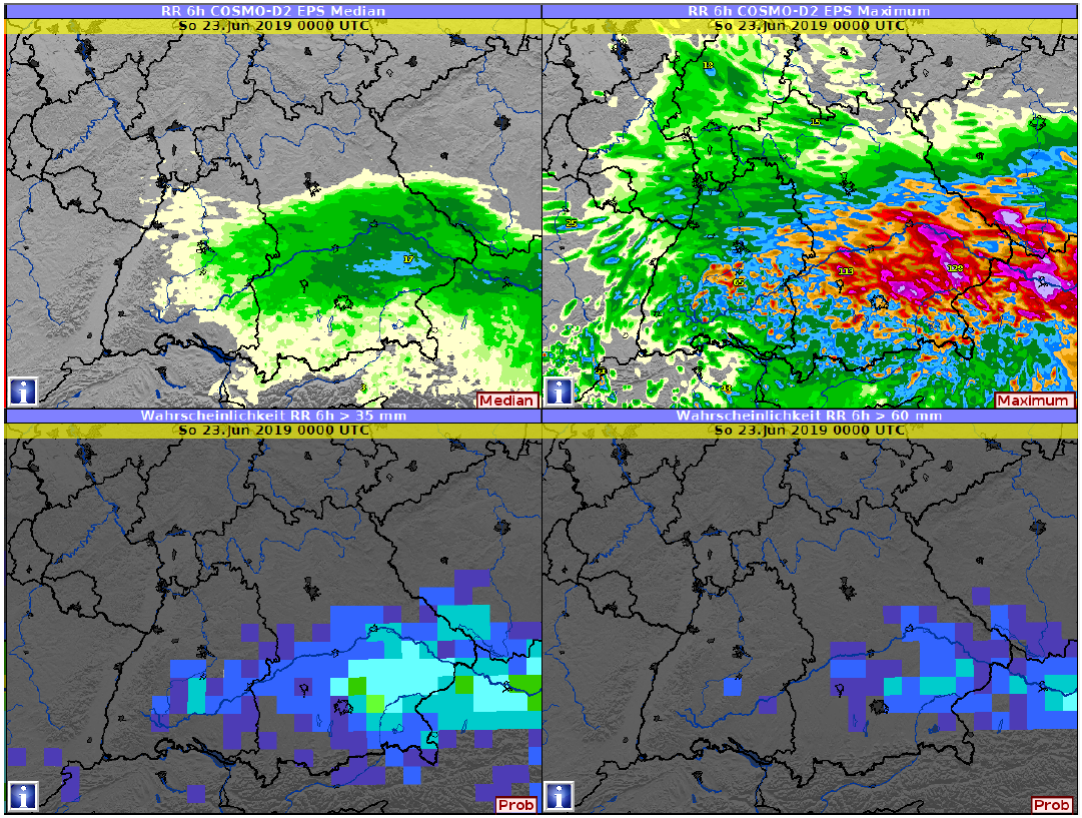
COSMO-D2-EPS

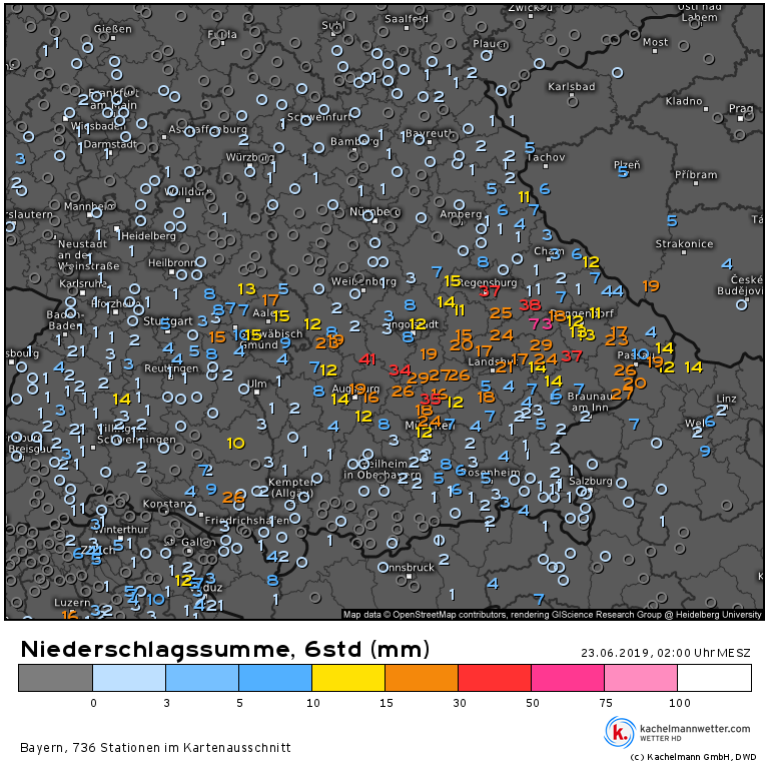




DWD, T. Schumann et al.

COSMO-D2-EPS

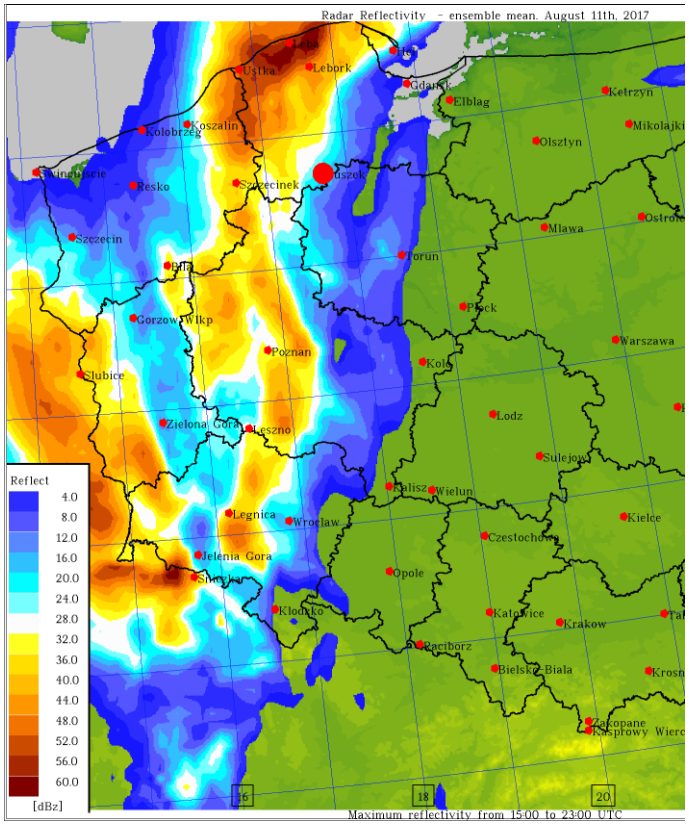




DWD, T. Schumann et al.

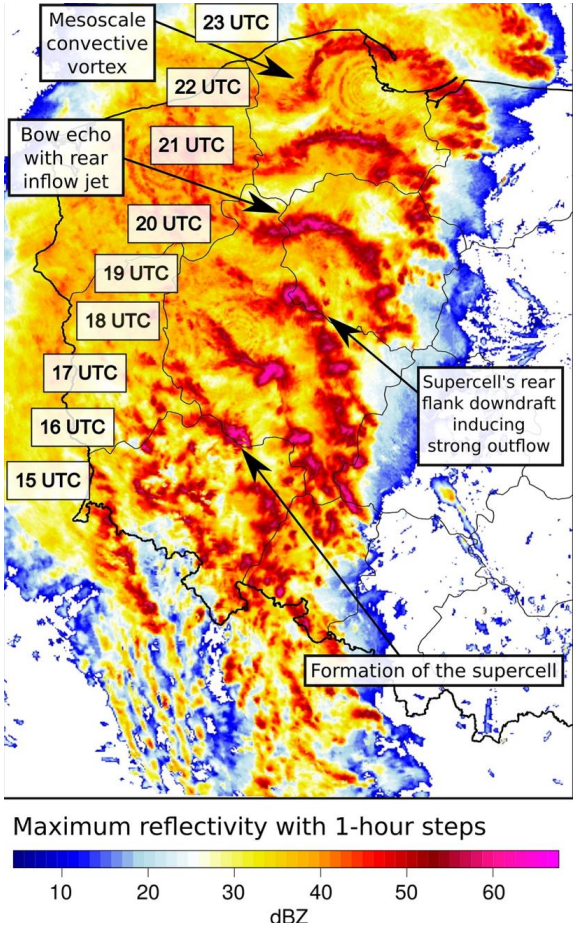
TLE-MVE for an High Impact Weather event

radar reflectivity forecast
ensemble mean



A. Mazur, G. Duniec, IMGW

observed reflectivity
Polish radar network

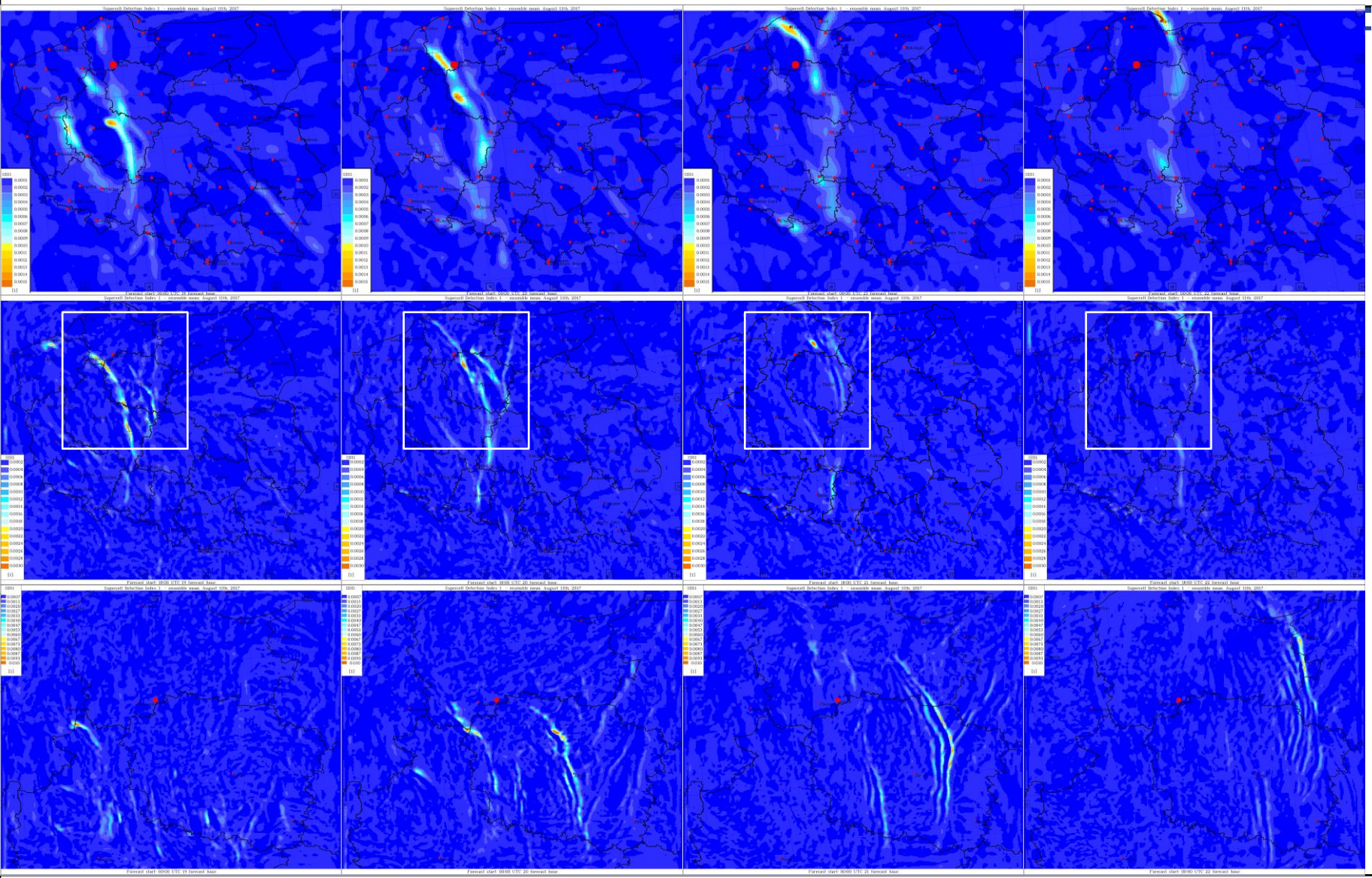


from Taszarek *et al.* (2019):
Derecho Evolving from a Mesocyclone

Single case – HIW event from 7 to 0.7km

DWD

Supercell Detection Index (SDI), ensemble mean

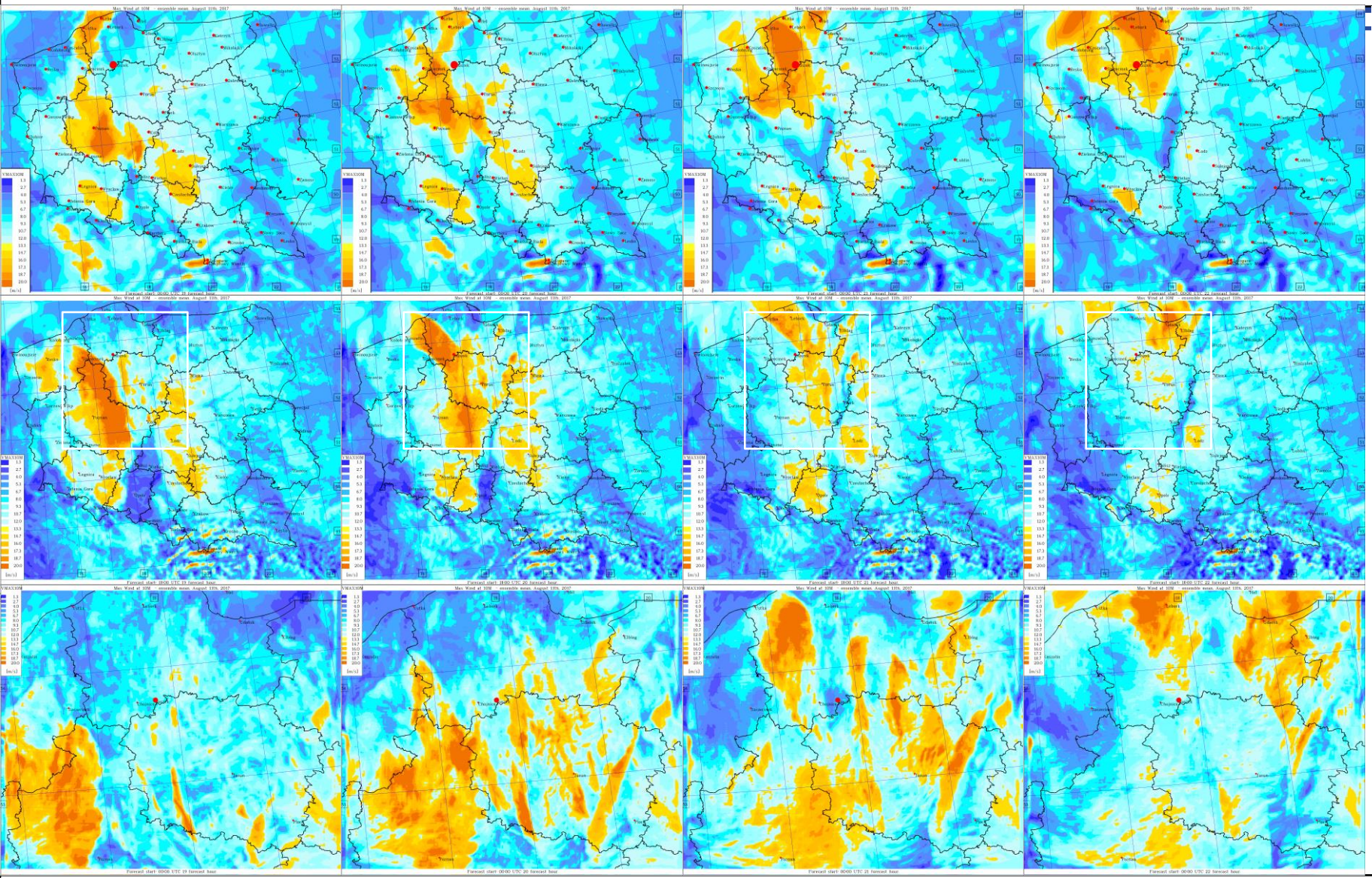


Suszek, August 11th, 2017

Single case – HIW event from 7 to 0.7km

DWD

VMAX, ensemble mean



Suszek, August 11th, 2017

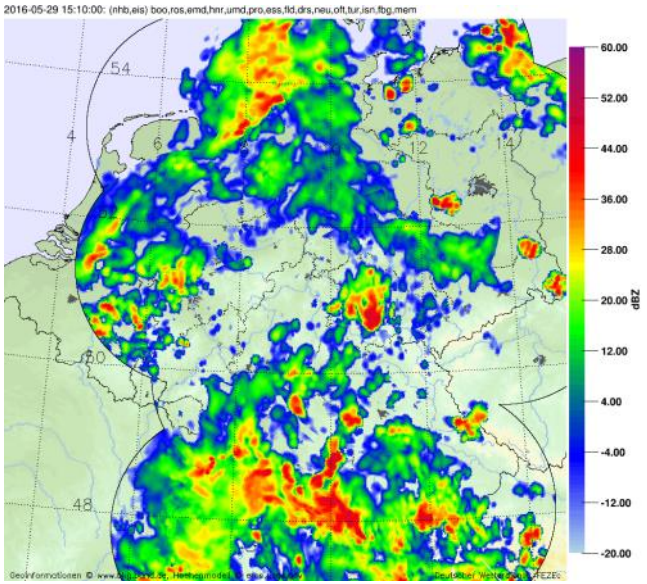
Ensemble verification

- Developer perspective:
  - Reliability
  - Spread/skill assessment
  - Usually for continuous variables (temperature, humidity, wind)
- User perspective
  - Catch the event
  - Spread/skill relation does not manifest itself on a single day
  - Needed for products (thunderstorm precursors, fog conditions, …)
  - verification for high impact weather
    - PP AWARE

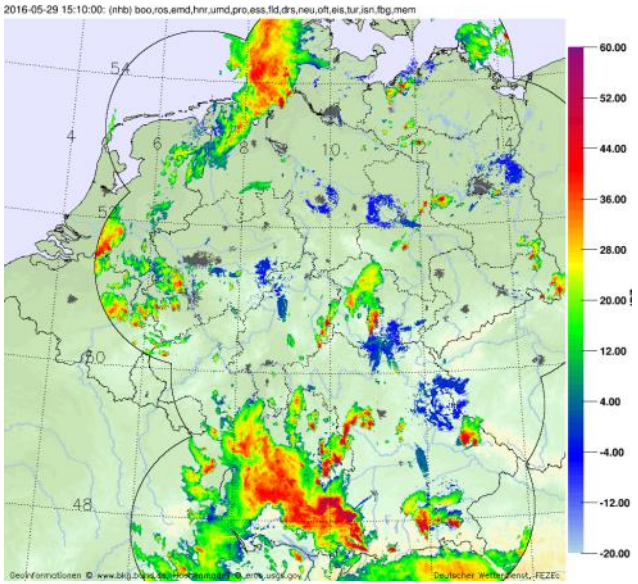
Verification of simulated reflectivities

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Predicted radar reflectivity



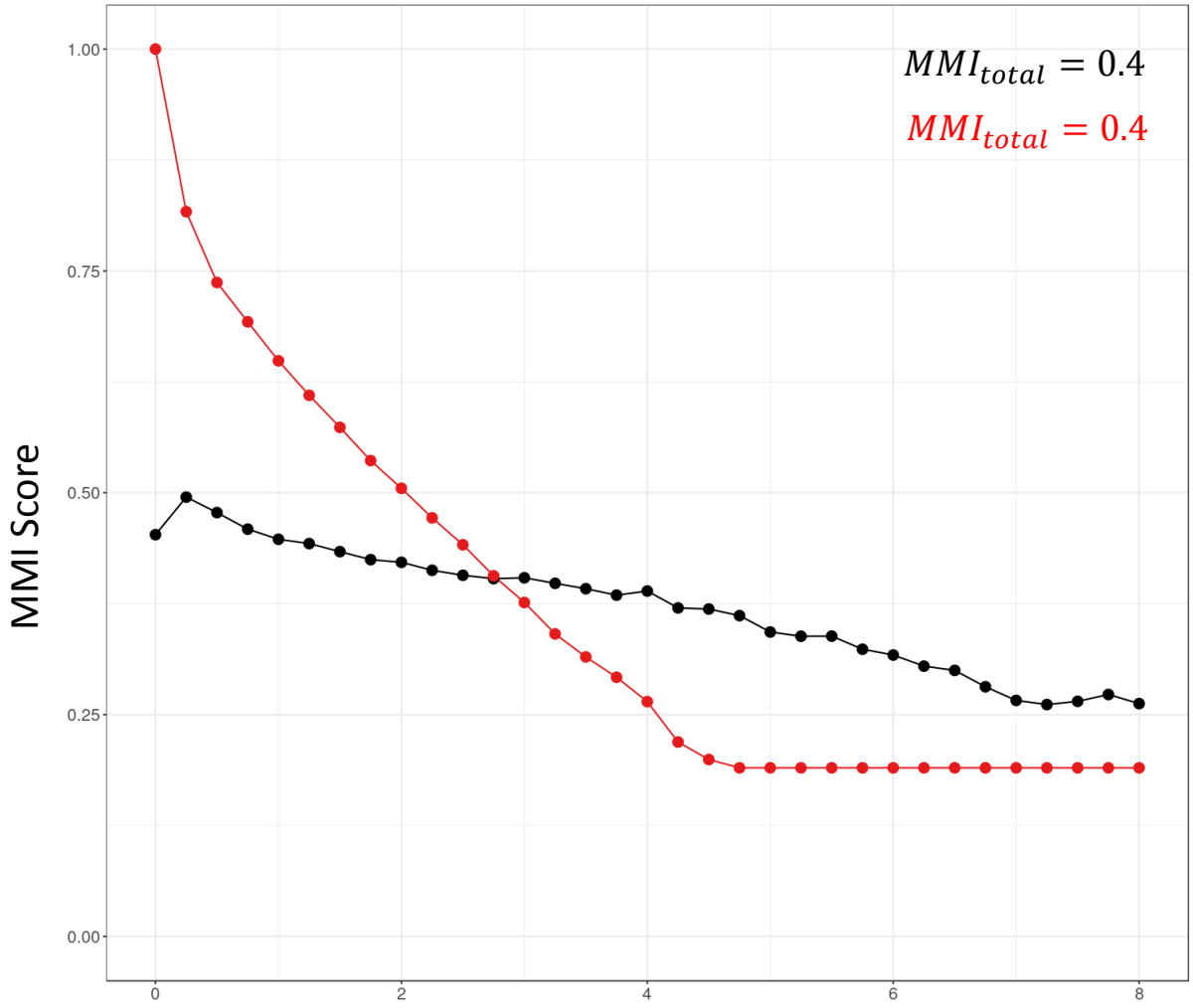
Observed radar reflectivity



- model results in observation space (dBZ)
- comparing apples with apples (?)

SINFONY

Verification of simulated reflectivities



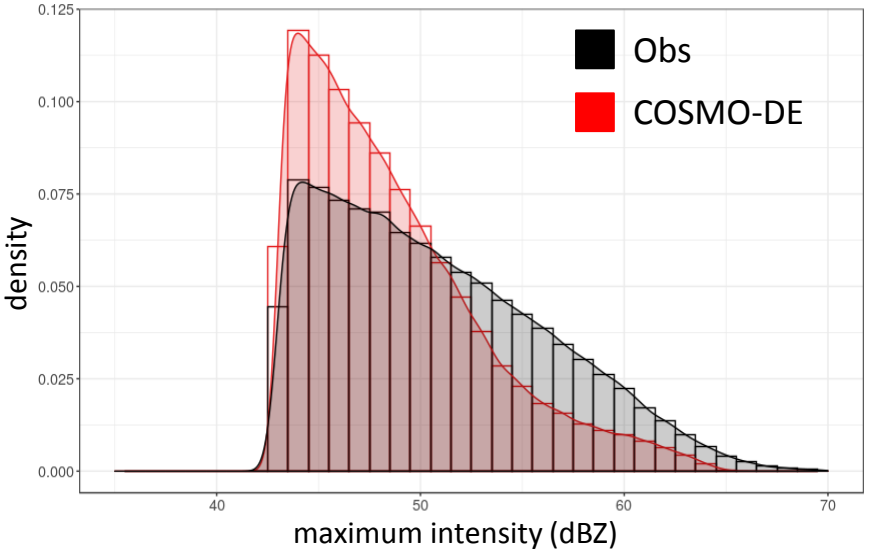
M. Hoff, F. Fundel, DWD

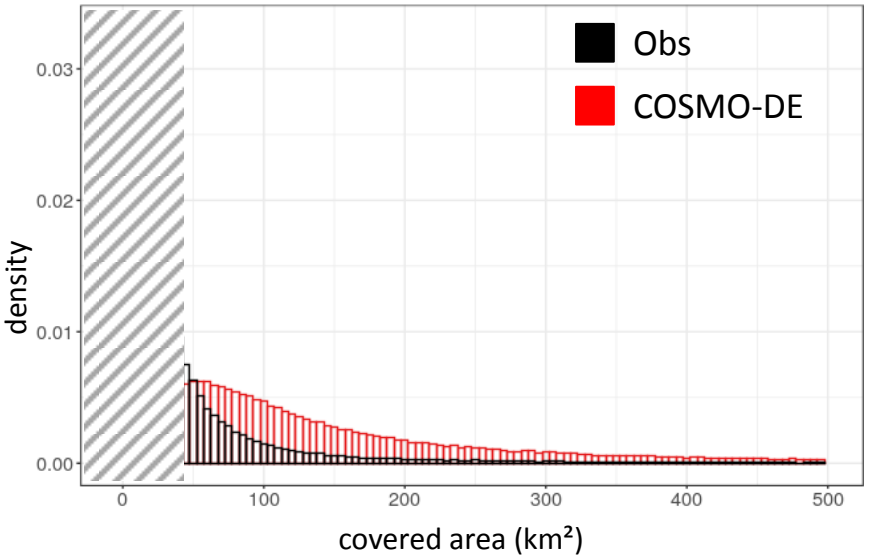
Verification of simulated reflectivities

- Is the comparison between observation and model forecast really fair?

- too many small objects in observation
- too many large objects in forecast

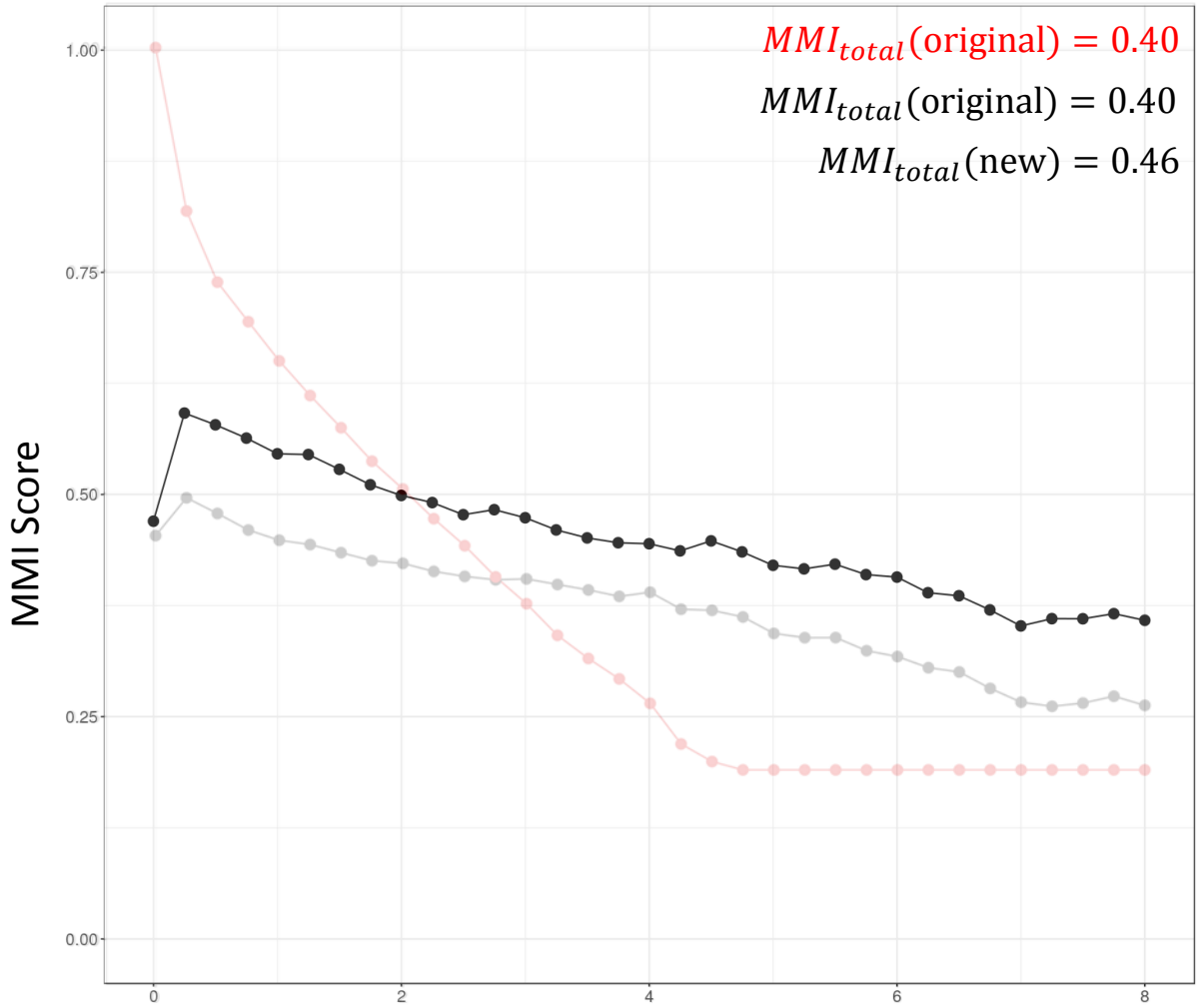
- experiment: What happens if we:
  - exclude features with area < 50km² (effective model resolution)
  - set 30 dBZ basic threshold to observed objects (instead of 35 dBZ) → results in larger observed objects





SINFONY M. Hoff, F. Fundel, DWD

Adapt the verification to predictable scale and biases



M. Hoff, F. Fundel, DWD

Lessons learned?

- The ensemble should provide a **forecast of the forecast error**
- A good model is **necessary** for creating a good ensemble, but not **sufficient**!
- A good ensemble **perturbation strategy** is needed, i.e. a good description of the model uncertainty and of the initial and boundary condition uncertainty
- The ensemble (should) describe the forecast error and not the systematic model error, but only the random component
- Increase of spread when the model systematic error is removed (easy to see under specific conditions, where the model systematic error is highlighted)
- ENS and DA: two problems with **almost** the same solution?
- Products and verification should be adapted to the **predictable** quantities