ML for Earth System Modelling and Analytics workshop 2021

04.05.2021

A ML-based post-processing of 2-m air temperature model output – a multi-model approach

Authors: Gabriel Stachura, Bogdan Bochenek, Marcin Kolonko, Piotr Sekuła, Małgorzata Szczęch-Gajewska

Affiliation:

Institute of Meteorology and Water Management
National Research Institute

Institute of Meteorology and Water Management
National Research Institute

Main issue

- several limited area models at the Polish NWS
- Which one is the best? It depends... (on place, season, weather parameter e.t.c.)
- Forecasters reconsile model outputs based on their knowledge and experience

The goal is to:

✓ improve accuracy of air temperature numerical forecast by reducing systematic error (bias)

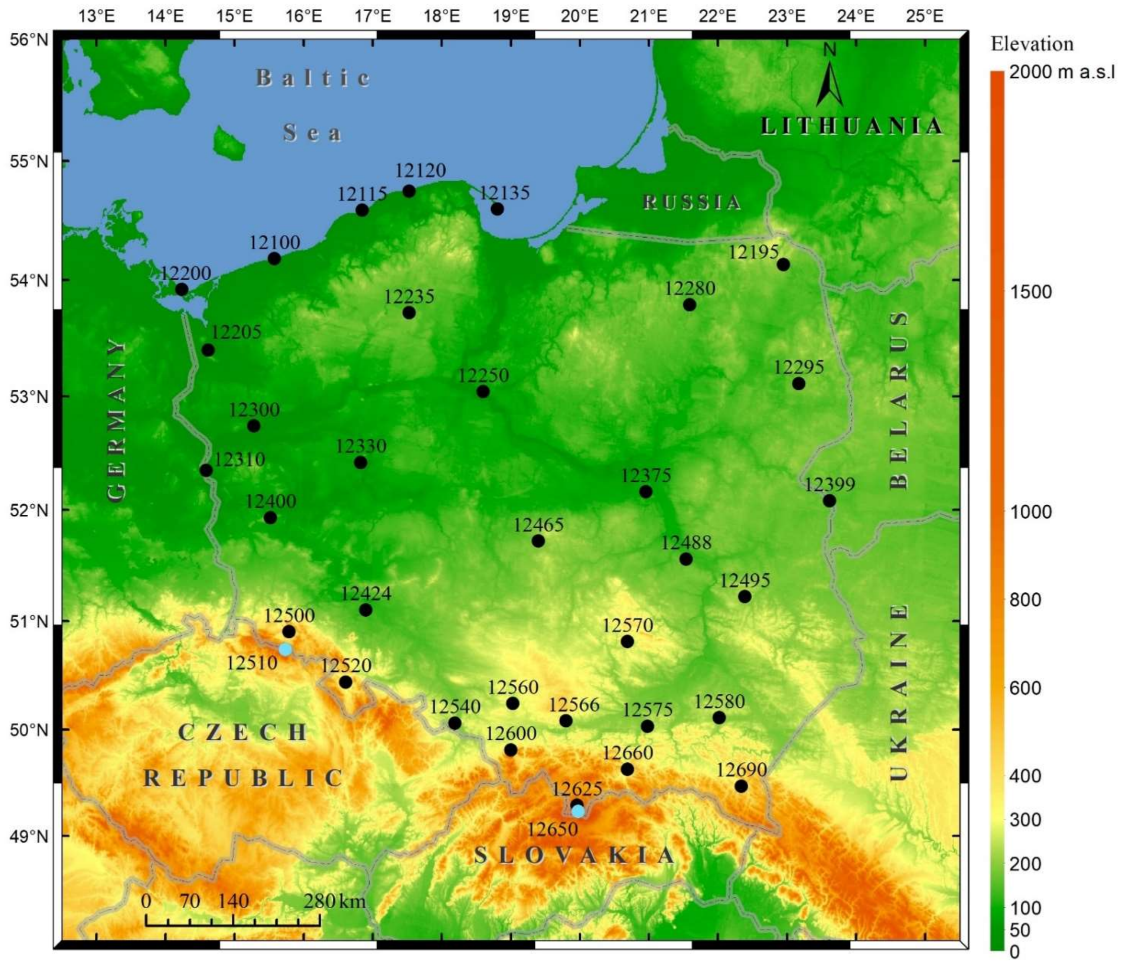
✓ Support forecasters in a decision-making process by an automated and objective tool

$T2m_A$ $T2m_B$ $T2m_C$ ⟷ $T2m_{OBS}$ ⟶ $T2m_{pred}$

Details of the experiment

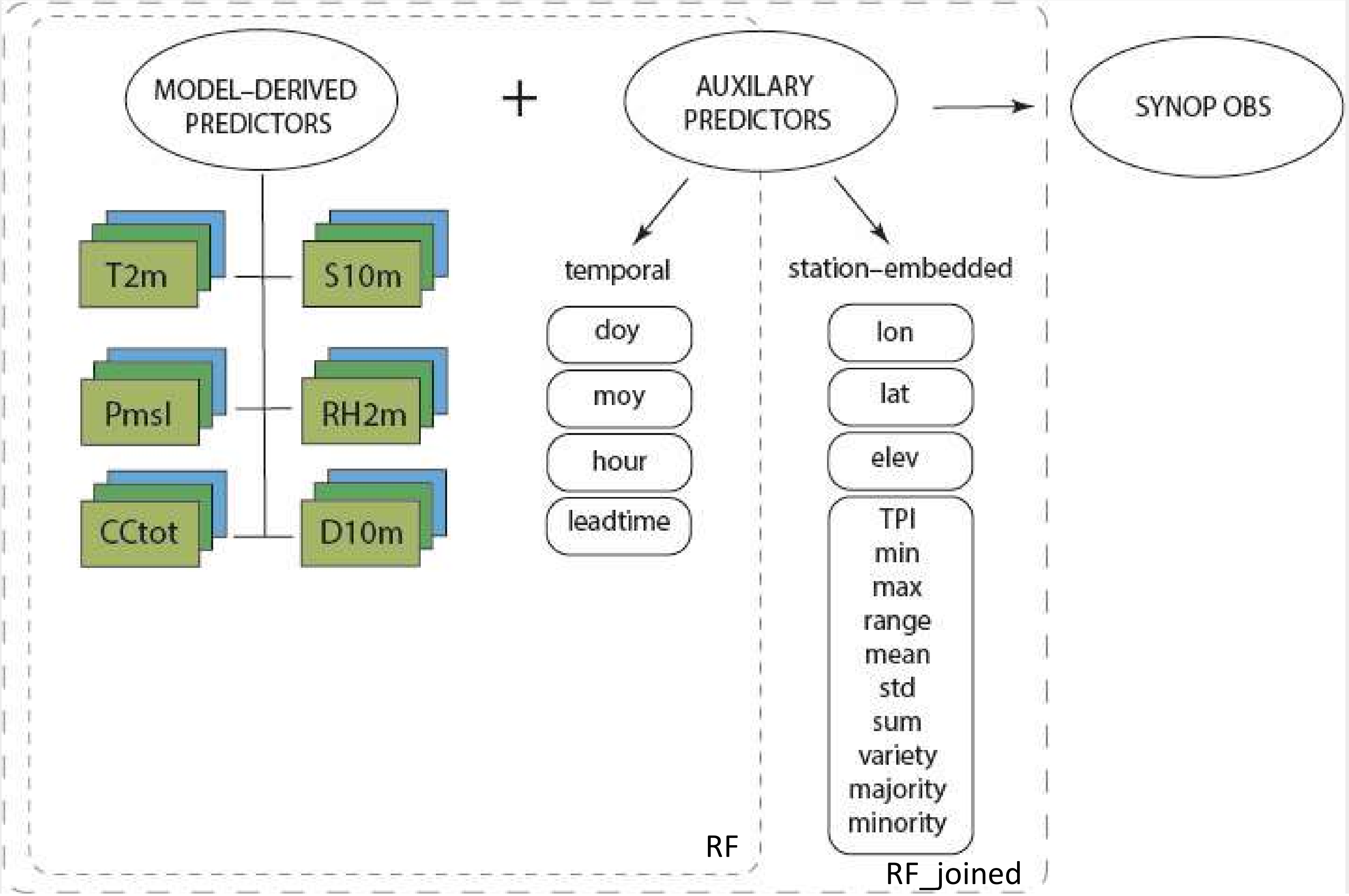
Institute of Meteorology and Water Management
National Research Institute

- 2-m air temperature
- Forecasts from three LAMs:
  - ALARO (4 km)
  - AROME (2 km)
  - COSMO (7 km)
- Forecast length: up to 30h
- 00 UTC run
- Temporal resolution: 3h
- ML method: RF, 1D regression, R software
- Training period: 2018-2019
- Test period: 2020
- Domain: 35 synoptic stations in Poland
- 4800 cases for every station



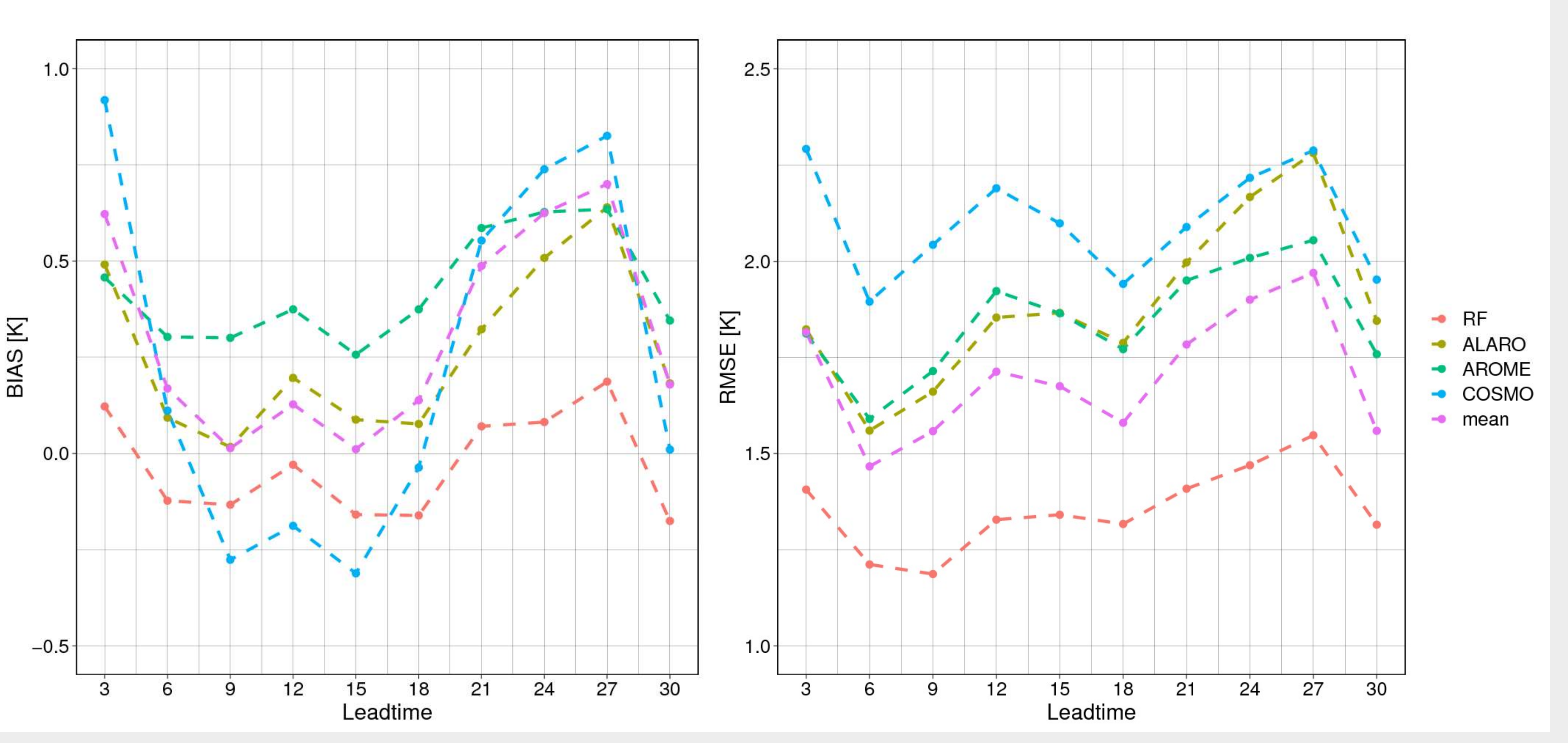
Training scheme

Institute of Meteorology and Water Management
National Research Institute



Results: temporal variability

Institute of Meteorology and Water Management
National Research Institute

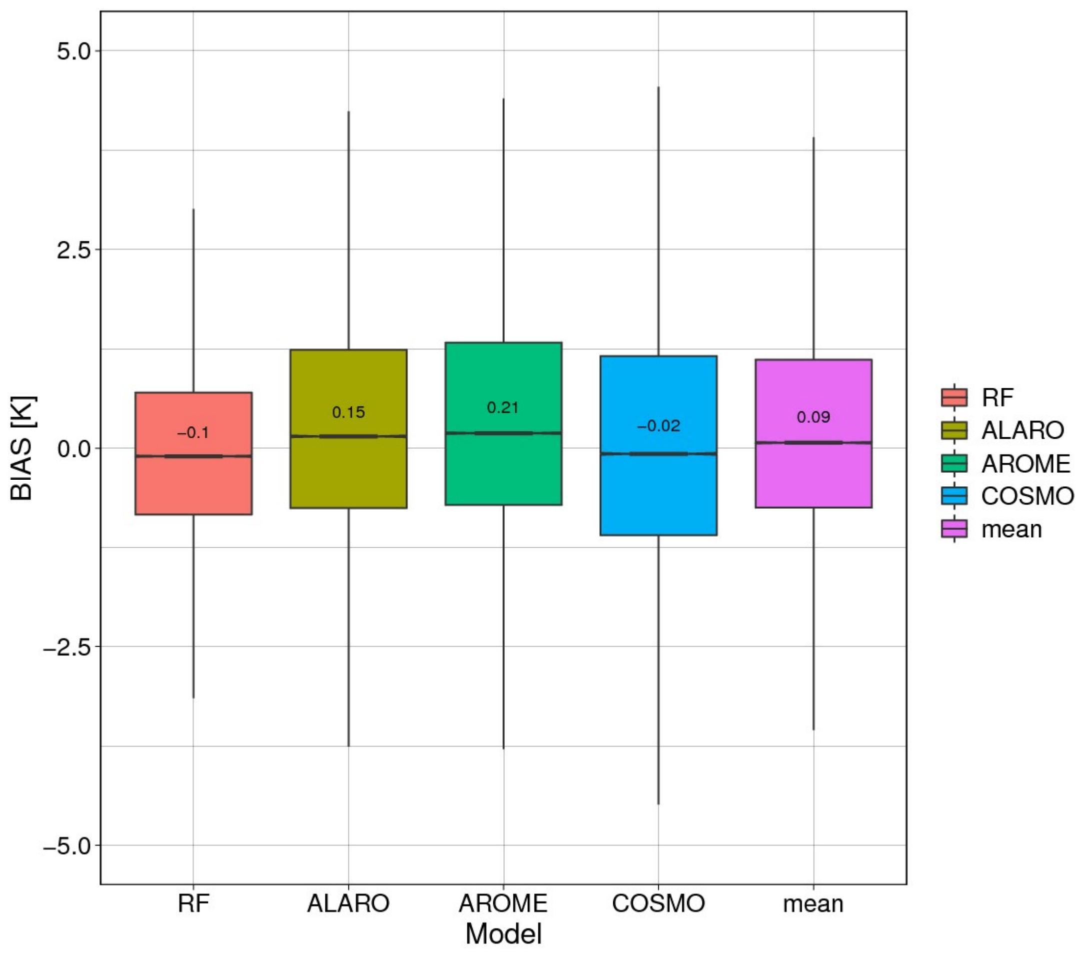


Mean skill score $= \frac{RMSE_{mean} - RMSE_{RF}}{RMSE_{mean}} \cdot 100\% = 20.4\%$

(26.6% with reference to AROME)

Results: bias distribution

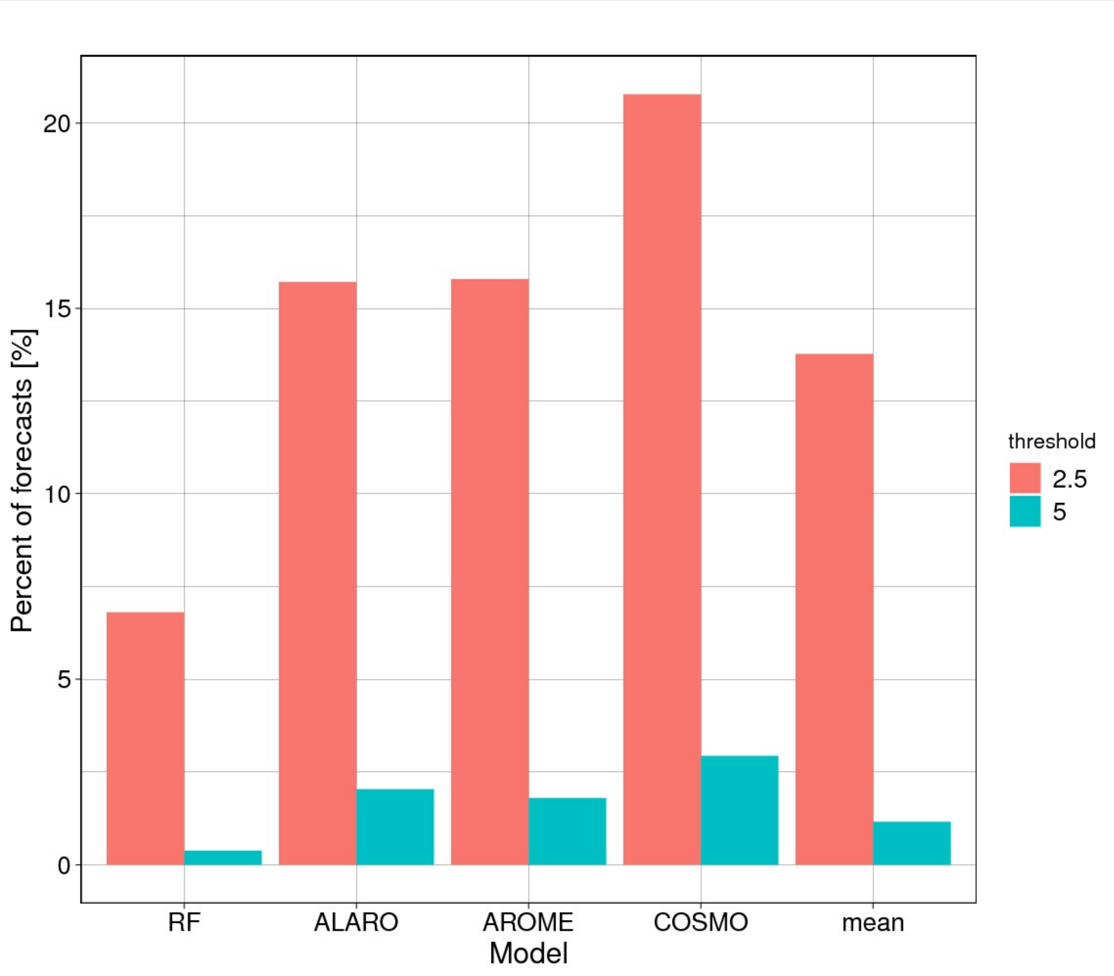
Institute of Meteorology and Water Management
National Research Institute



- The lowest interquartile range
- The lowest spread after removing outliers

Results: accuracy

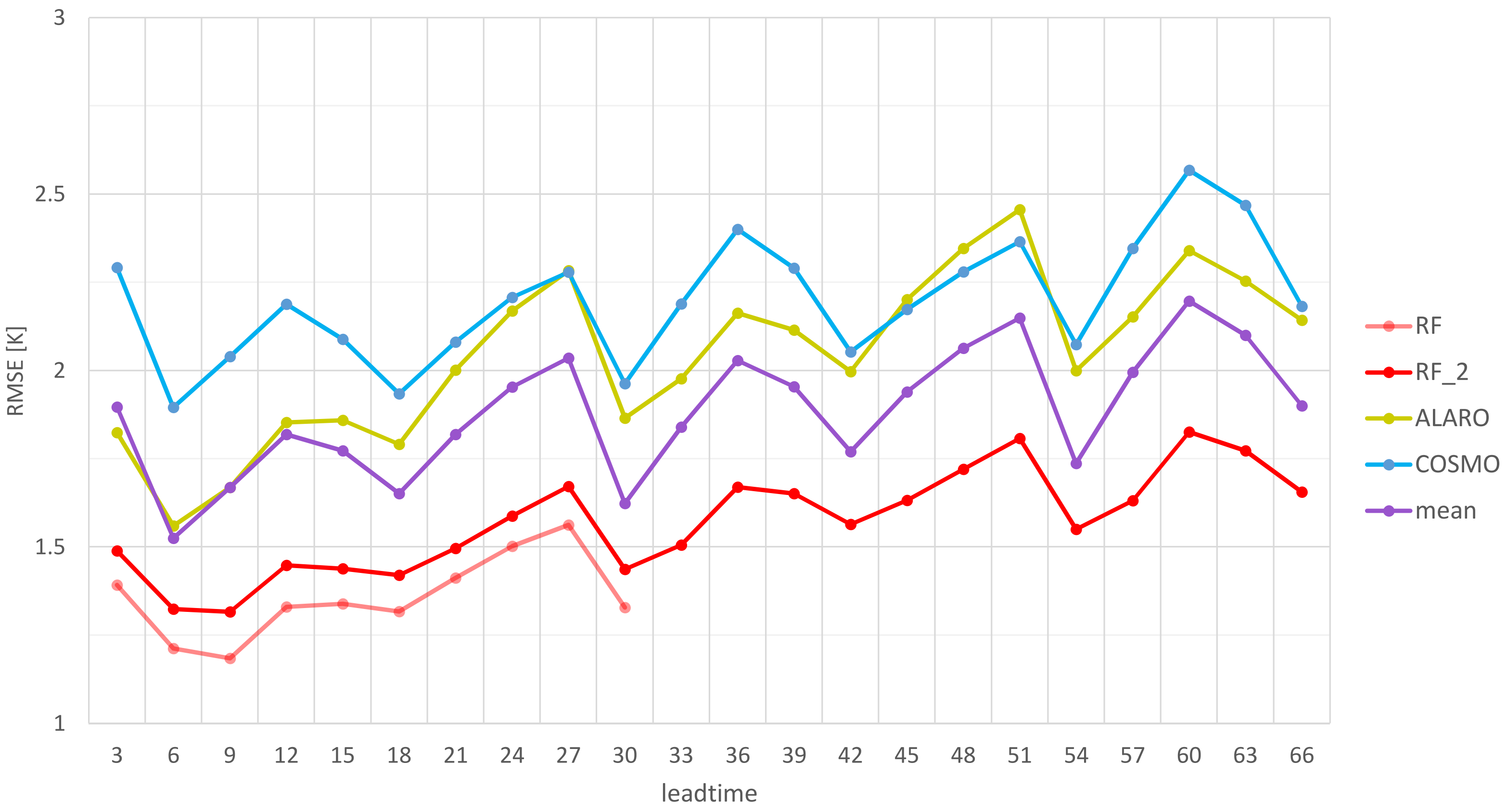
Institute of Meteorology and Water Management
National Research Institute



- Two thresholds for forecasts error: 2.5 and 5°C
- RF makes mistakes twice less often then other methods

Results: case study

Institute of Meteorology and Water Management
National Research Institute



- AROME was removed from the database => forecast length up to 66h
- Slightly worse results within 30h leadtime, but still better than other methods
- RF's trend seem to be similar to other's
- About 2 days gain in RMSE

Conclusions

Institute of Meteorology and Water Management
National Research Institute

- RF remarkably reduces systematic error of the operational models and improves forecast accuracy
- The advantage of RF is particularly distinct in case of RMSE => RF corrects bigger biasess better than the smaller ones
- For longer leadtimes, the error grows roughly paralel to input model's errors
- Training all stations at once enables potential operational use of our method
- To examine: performance on a spatially continuous domain?