46th EWGLAM and 31th SRNWP Meeting, 30/09 – 03/10/2024 Prague & online

METEO IMGW-PIB meteo.imgw.pl

LACE nwp central europe

ACCORD A Consortium for COnvection-scale modelling Research and Development

NWP in Poland

Piotr Sekuła, Marcin Kolonko, Jadwiga Róg, Gabriel Stachura, Natalia Szopa, Małgorzata Szczęch-Gajewska, Bogdan Bochenek

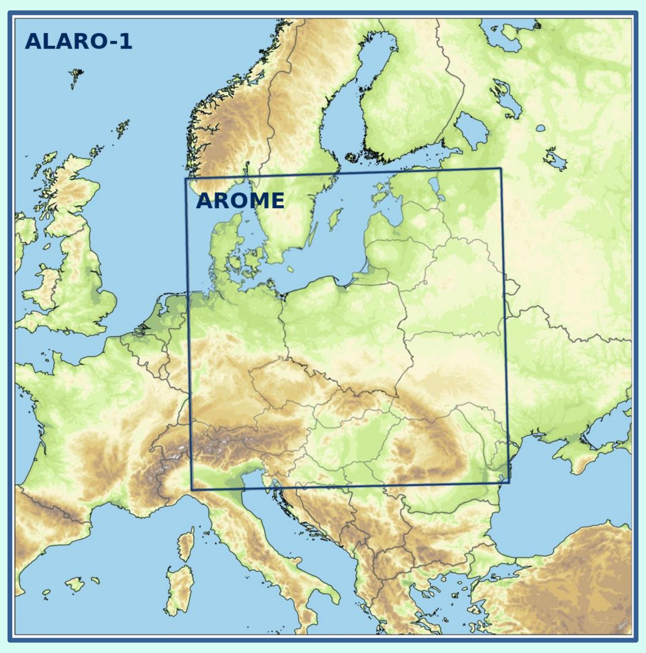
Operational

ALARO-v1B NH (CY43T2) operational domain:

4km horizontal resolution, **789x789** grid points, **70** vertical model levels, on a Lambert projection with **3h** coupling frequency and **1h** output, coupling zone 16 points; Runs 4 times per day (00,06,12 and 18) with **72** hours forecast range; LBC from ARPEGE with 9.4km horizontal resolution; Time step ~160s. In pre-operational mode CANARI surface assimilation with 6h cycling.

AROME (CY43T2) operational domain:

2km horizontal resolution, **799x799** grid points, **70** vertical model levels on a Lambert projection with **3h** coupling frequency and **1h** output. **4** runs per day (00, 06, 12 and 18UTC) with **30** hours forecast range; LBC from ALARO-1 4km; output every 1h – for LEADS system; 10min output for INCA Nowcasting System.



Computational domains of ALARO-1 (4.0km horizontal resolution) and AROME (2.0km horizontal resolution) nested models.

Operational machine characteristics

Cluster of HP BL460c_GEN8 servers connected with Infiniband network, OS Scientific Linux 6, Intel Xeon E5-2690 processors – with maximum 1552 cores (97 nodes with 16 cores each), each core RAM 128 GB, disc array – 64 TB.

Ensemble forecasts using AI global weather models

Model information

FourCastNet is a data-driven global weather forecasting model. It was trained on ERA5 reanalysis data, spanning years 1979-2015. It can forecast 20 atmospheric variables, 5 of which are on the surface level. Timestep between each forecast is 6 hours, with a 0.25° spatial resolution, and the result is a 720x1440 grid point map of predicted values for each variable.

The model used in this project is a modified version of FourCastNet, installed on the ACK Cyfronet computing center supercomputer Athena. It creates an ensemble forecast, and the number of ensembles rises with the number of timesteps. For *n* timesteps, the final ensemble count is 2 to the *n*-th power.

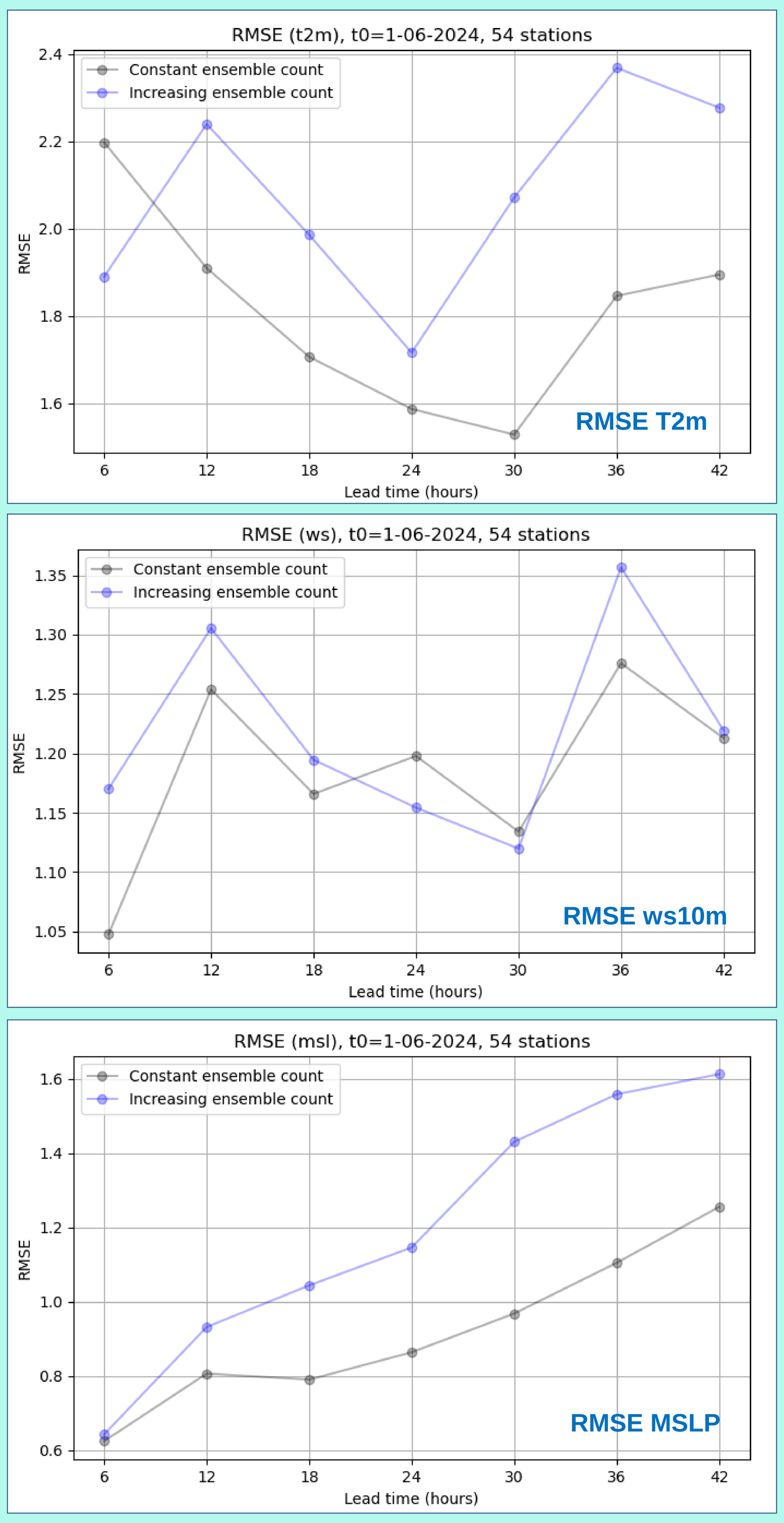
Methodology and data

The aim of the experiment was to generate AI-driven ensemble forecasts using 2 different approaches, and then compare their accuracy. First method has the ensemble count increase from 2 to 128, rising to the next power of 2 with every timestep. In the second method, the model maintains the same number of ensembles through the entire forecast (128 ensembles). In both cases the initial conditions are perturbed with a Gaussian noise function.

Only surface variables were considered: air temperature at 2m height, mean pressure at sea level, and wind speed. The testing period was June 2024, with the model being run each day at 00:00 UTC for 7 timesteps (42 hours). Additional data used were SYNOP observatios from 54 meteorological stations in Poland. This data were used to calculate forecasts RMSE for all stations, and the error was compared for both methods.

Results

Air temperature: Variable with the highest RMSE values. The mean RMSE values initially decrease with time. For the increasing ensemble count, 4th timestep has lowest RMSE, which corresponds to 00:00 UTC the next day of the forecast. For the constant ensemble count it's the 5th timestep, 06:00 UTC. We speculate that the variation in RMSE is tied to the amplitude of temperature during the diurnal cycle, and daytime temperature fluctuating more than nighttime one. The error for the 1st step with 2 ensembles is 0.4°C lower than with 128 ensembles, but otherwise the model with the constant ensemble count seems to perform better.



Root mean square error for various Polish synoptic stations, for the following variables: Temperature at 2m [°C], mean pressure at sea level [hPa], wind speed [m/s].

Wind speed: Calculated from zonal + meridional components. The constant count method performs better, but not as visibly as for other variables. The increasing method's error is smaller on 4th and 5th timestep. The lack of trend may be tied to the natural variability of wind speed.

Pressure at sea level: RMSE is, more predictably, rising with every step. Along with the ensemble count increasing from 2 to 128 consistently worse results appeared, even though there is almost no difference between the methods at the beginning. The biggest difference (over 0.4hPa) is on the 5th timestep. It's the element with the lowest RMSE overall, but the error steadily rises over time.

Verification of ALARO-1 CY46T1 and impact test of 1h LBC

CY46T1 tests runs

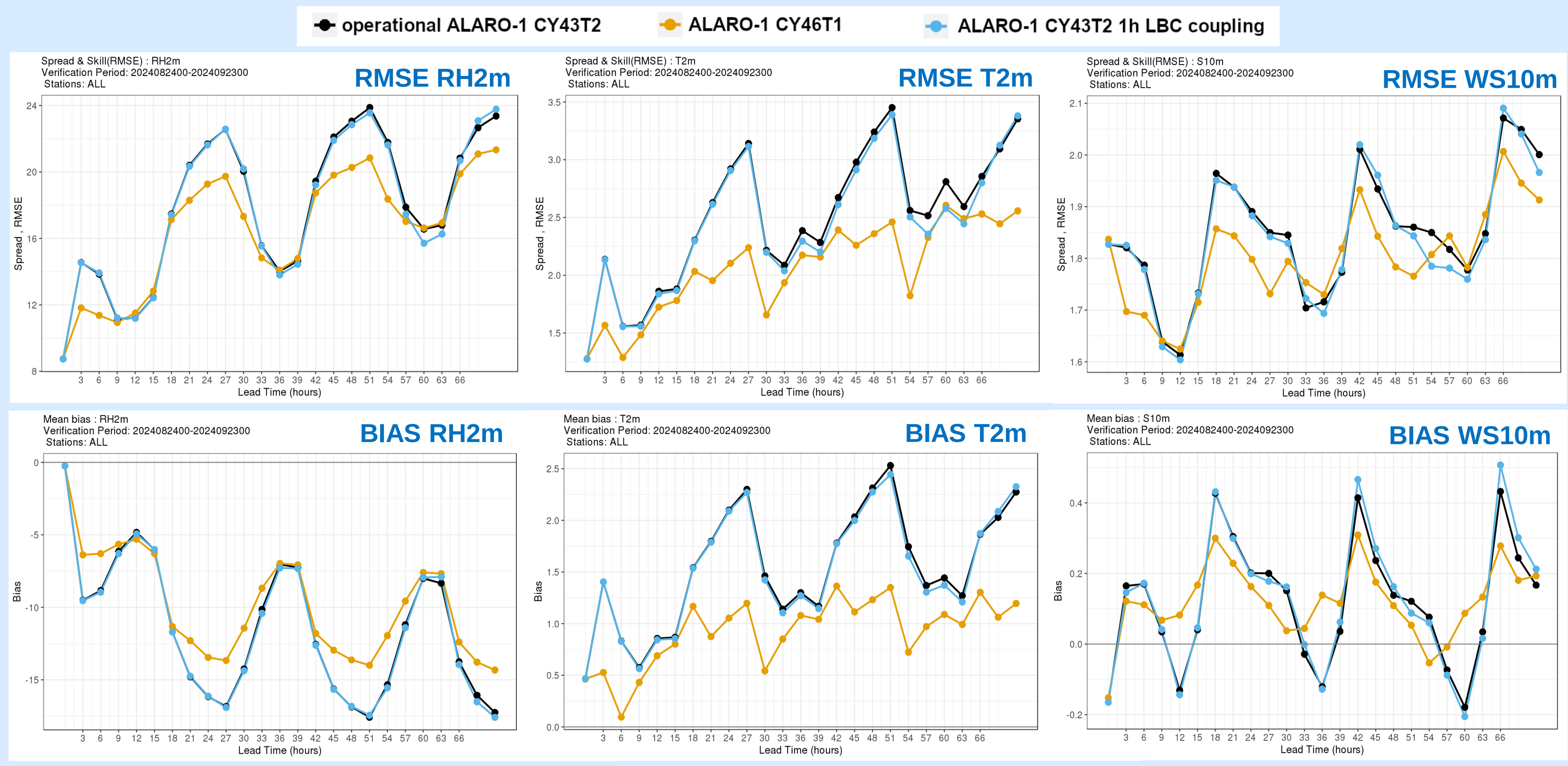
Preoperational tests with CY46T1 export version runs daily for ALARO-1 CMC. Start of the test was in February 2024, but till May some corrections were incorporated into this export version. These were four packages of code changes developed by Czech LACE team in Prague.

In general, tests show significant improvement of forecast comparing to current operational version (CY43TTbf11). One month (2024.08.24-20224.09.23) verification results comparison of chandged ALARO-1 CY46T1 (orange lines) with with respect to operational run (black lines) for 2m temperature (T2m), relative humidity (RH2m) and 10m wind speed (WS10m) are presented on the common graphs below.

CY43T2 1hour coupling tests runs

Parallel tests of new LBC from ARPEGE with 1-hour coupling frequency are running daily for ALARO-1 CMC CY43T2bf11 (currently operational version of model) almost from begining of 2024.

The first results of the verification, performed on the fly, do not show any significant improvment of forecasts when running the same model with 1h LBC frequency compared to 3h (operational) for overall scores. Further studies of 1h coupling frequency for individual cases are planned. On the graphs below verification of 1h coupling forecasts represent blue lines.



Comparison of ISBA snow schemes in SURFEX

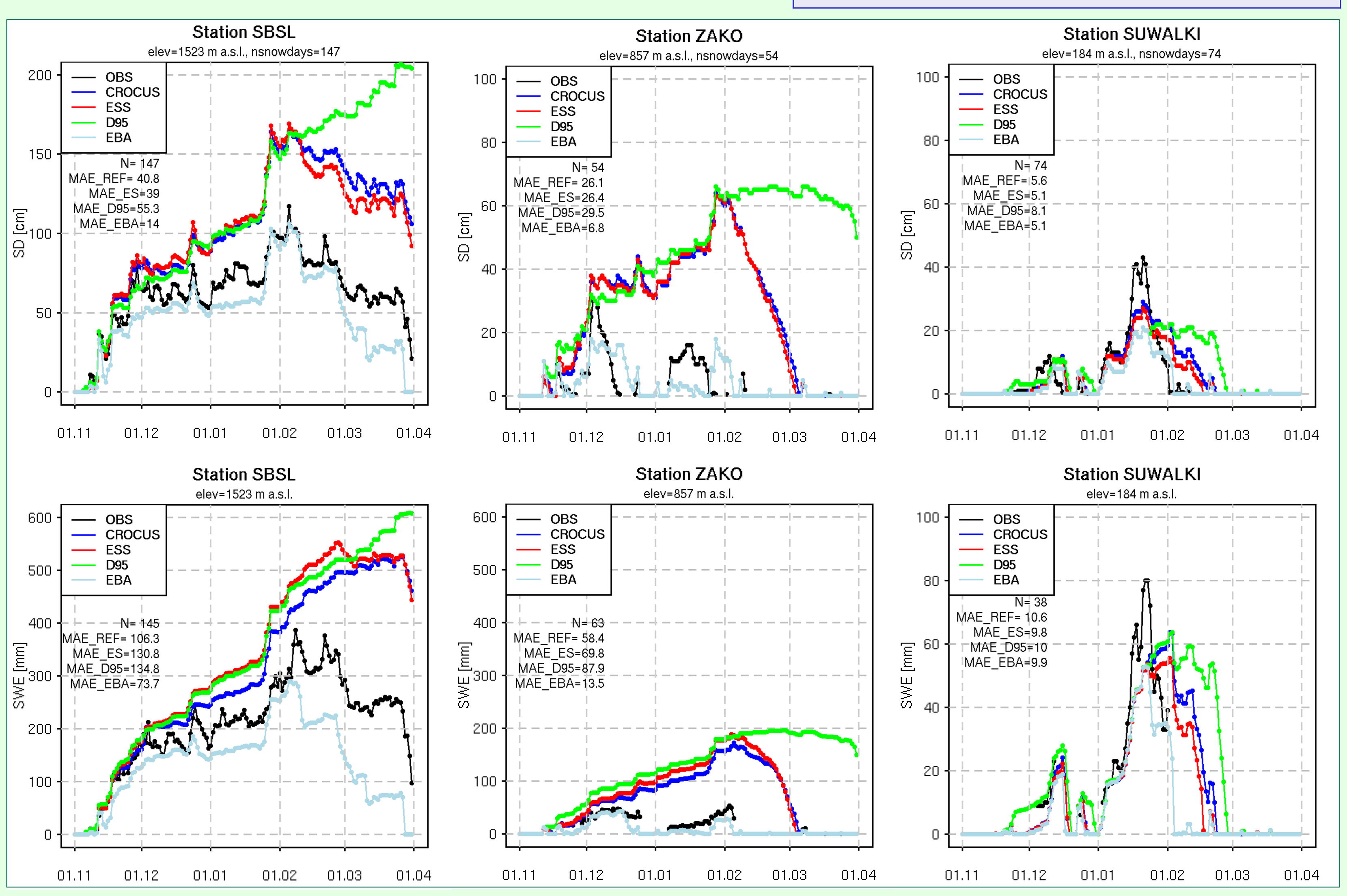
Methodology and data

Four snow schemes available in SURFEX 8.1 have been compared and verified regarding snow properties. These are two one-layered, composite schemes: **D95** and **EBA** and two multi-layered schemes: **Explicit Snow Scheme** (ESS) and **CROCUS**. Basic setup of the experiments is shown in the Table 1. The only parameters changed in the namelist was CSNOW and CISBA.

Comparison spans the latest winter season (from November 2023 to March 2024). Two snow variables were evaluated: snow depth (SD) and snow water equivalent (SWE). On the charts below you can see the results from one high-mountain station (SBSL, 1523 m a.s.l.), one mountian-valley station (ZAKO, 857 a.s.l.) and one lowland station (SUWALKI, 184 m a.s.l.). The metrics on the charts were calculated only for cases when snow cover was present.

Table 1. SURFEX setup used in the experiment.

forcing model	AROME (2.5 km)
forcing frequency	1 h
forecast length	24 h
forecast timestep	900 s
NPATCH	1
CISBA	DIF / 2-L
TEB	off
CSNOW	CRO / 3-L / D95 / EBA
data assimilation	off



Forecasted and observed snow depth (upper row) and snow water equivalent (bottom row) in winter season 2023/2024.

Conclusions

In the period of snow accumulation, D95, ESS and CROCUS peform roughly similar. Their positive bias gradually increases. EBA is the only scheme which underestimates snow cover, however, this uderestimation is also small. During melting episodes, EBA responds the fastest and corresponds to observations the best. Other schemes tend to respond with a delay and reduce snow cover insufficiently. This tendency is the most apparent in spring. In D95 scheme, melting is extremely delayed (around one month in the lowlands and around two months in the mountains). ESS and CROCUS are very similar regarding snow depth, however, it is interesting to notice that SWE is considerably larger in ESS. In spring, ESS melts snow faster than CROCUS. Overall metrics favour EBA as the most accurate snow scheme.