4th ACCORD All Staff Workshop, 15 – 19 April 2024, Norrköping & hybrid

METEO IMGW-PIB
meteo.imgw.pl

NWP in Poland

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

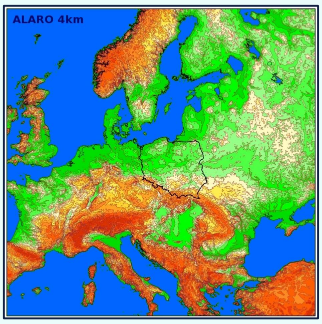
Operational

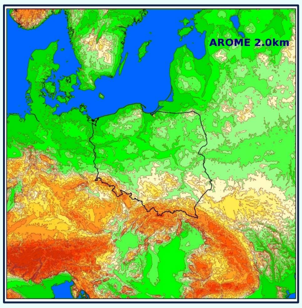
ALARO-v1B NH (CY43T2)

Operational Domain:

E040 domain:

4.0 km horizontal resolution, 789x789 grid points, 70 vertical model levels on a Lambert projection with 3h coupling frequency and 1h output, coupling zone with 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 150s.





AROME (CY43T2)

Operational Domain:

P020 domain:

2.0km horizontal resolution, 799x799 grid points, 70 vertical model levels on a Lambert projection with 1h coupling frequency and 1 hour output 4 runs per day (00, 06, 12 and 18UTC) with 30 hours forecast range; Time step 50s; LBC from ALARO-1 4.0km; GRIB format, every 1h – for LEADS system; 10min output for INCA Nowcasting System.

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.

Comparison of ISBA snow schemes in SURFEX

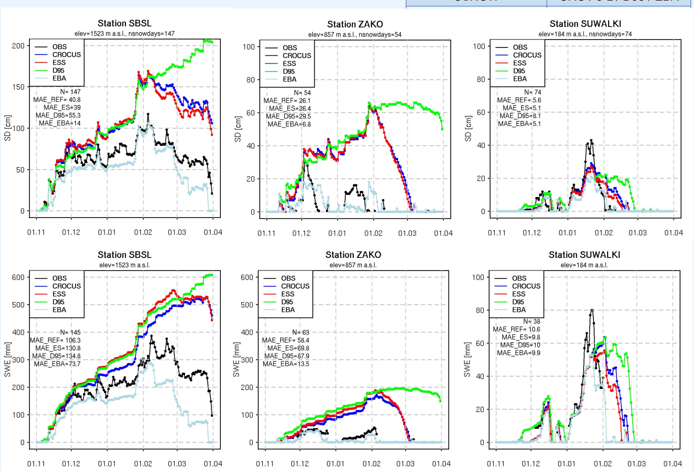
Methodology and data

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 results from the results from one high-mountain station (SBSL, 1523 m a.s.l.), one mountain-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.

	value
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



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.

Vine copula application to ensemble postprocessing

Mathematical background

A *d*-dimensional copula is a multivariate distribution function on $[0,1]^d$ with uniformly distributed marginal distribution functions. Fundamental Sklar's theorem states that for a *d*-dimensional cumulative distribution function, there exists a *copula* function denoted by *C*, such that $F(x_1(t),..,x_d(t)) = C(F_1(x_1(t)),\dots,F_d(x_d(t)))$, where F is a joint cumulative distribution function and $F_1,..F_d$ are marginals. This theorem allows to separate univariate margins from the dependence structure.

Method

The goal of the copula-based method is the error mitigation of the temperature forecast given by the ALARO model. We construct a copula that contains both the information about the correlations between variables affecting the forecast error and their individual probability distributions. From a copula-given conditional probability distribution we are able to obtain a sample of pseudo-observations. We aim to check whether the choice of different conditioning variables has a significant effect on the correct fit of the model to pre-existing real data. We then calculate the average of these generated forecast errors, which we then add to the ALARO model's temperature forecast and check how much the corrected forecast is better than the original forecast using RMSE.

Indicator	Description of the conditioning variables
a	AROME model forecast for the current day
b	COSMO model forecast for the current day
c	Forecast error of the ALARO model on the previous day
d	Value of observed temperature at 00UTC
e	Forecast error of the AROME model on the previous day
f	Forecast error of the COSMO model on the previous day
g	Forecast error of the AROME model on the current day
h	Forecast error of the COSMO model on the current day
i	Difference between the forecast on the previous day and the current day of the ALARO model
j	Difference between the previous day's relative humidity forecast and the current day's ALARO model forecast

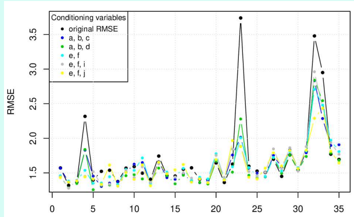
Data

Considered data will include forecasts of air temperature values at 2m above ground level of NWP models at 12UTC (forecast initialised at 00UTC) for 35 Polish synoptic stations. The dataset was divided into a training set (forecasts from January 1, 2019 to December 31, 2019) and a test set (forecasts from January 1, 2020 to December 31, 2020).

Results

We see that the chosen method introduces a slight correction in the temperature prediction of the ALARO model. We verify the results by observing the percentage change in the root mean squared error (RMSE) of the ALARO model's temperature prediction. RMSE is an indicator that measures the average difference between the model's forecast values and actual values.

The largest improvement due to using that method was for stations with high systematic errors – mountain and seashore ones. Significant reductions in RMSE are observed at the station located at Śnieżka, where the error was reduced by an average of 45%, and in Hel, by 28% per year. In Zakopane and on Kasprowy Wierch, by 18% and 17% respectively. The average largest improvement for all 35 stations was observed with the conditioning variables of temperature forecasts of different models (AROME and COSMO) and the value of observed temperature at 00UTC.



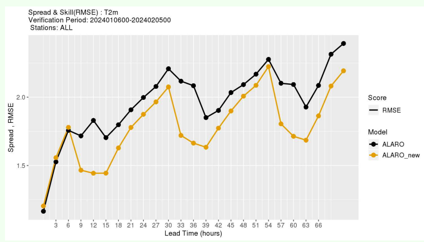
On the X axis – meteorological stations by ID, on Y-axis – RMSE for various sets of conditioning variables.

Conclusions

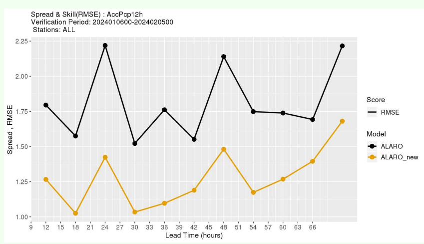
The vine copula method allows reduction of systematic errors and generation of synthetic multi-dimensional data obtained by models in the hindcast. We can generate a sample from copula probability distribution and thus save CPU which is critical to the ensemble forecasting.

Data assimilation in ALARO

In a test mode, we run surface assimilation using CANARI within CY43T2, with SYNOP data and partially automatic stations data (use of OPLACE database). Assimilation is running in 6h cycle. The training period was several months, to obtain the improvement of out assimilation results. The worst scores we observe for first hours of forecast (noise), the reason may come from lack of DFI (for some technical reason, on which we work). In general, our assimilation substantially improve the forecast, as one can see on verification scores graphs. The diagrams show scores (RMSE) for T2m and 12h accumulated precipitation. Similar results we obtain for other variables (RH2m, Pmsl, wind) available in HARP verification system.



Comparison of verification scores of ALARO CY43 forecasts (1-month period from 6th January to 5th February 2024) with DA (orange curve) and without DA (black curve)– 2m temperature.



Comparison of verification scores of ALARO CY43 forecasts ((1-month period from 6th January to 5th February 2024) with DA (orange curve) and without DA (black curve) – 12h accumulated precipitation.