Applications and Verification

Report

Prepared by:	Area Leader Doina-Simona Tașcu	
Period:	2023 (January - August)	
Date:	September 2023	

Introduction and background

The main activities within the “Applications and Verification” area are to collect information and identify LACE members needs and create an action plan to systematically manage, organize and control the changes in the software codes. Furthermore, to maintain a documentary fund which should help the NWP colleagues.

Goals

The primary goal is to develop and adapt/use different specific applications into user-friendly mode. Many tools and software products were developed along the years for meteorological parameters. These days, it is imperative to have easy to use applications, maybe to find and to identify the operational activities, to make a common way for saving time and manpower resources. Nowadays, it is important to make the applications easy to implement without too much cost and to make a common way for saving time, computer costs and manpower. It is a big challenge to identify and to merge all the beneficial technical approaches and applications for all countries.

Main R&D activities

It is well known that the verification of numerical model is an important tool in order to ensure a better use of the meteorological products for operational and research usage. The main activities are the following:

- HARP implementation and verification for deterministic and probabilistic forecasts
- HARP linked to OPLACE database
- Multiple verification methods
- Post-processing of model output
- RC-LACE github platform
- Database of cases
- Trainings

1 Action: HARP implementation and verification for deterministic and probabilistic forecasts

Implementation of Panelification tool in HARP (***Austria***)

In order to implement Panelification into HARP system, the following steps are necessary:

1. Setting up environment:

- conda (mainly for R version on virtual machine)
- renv (for R packages installed within R)

2. harpSpatial

- Added scores **rmse, Rpearson, FSS percentiles** to local harpSpatial
- **Distinguish** between **regridding** domain **and verification domain**
- Allow returning of regridded precipitation **fields & verification domain**
- Add **unit_factor** to assure same units of the fields

3. Script

- **Ranking** of scores (non-FSS, FSS, averaging of ranks):

0	NAN – black
1	FSS < fobs – red
2	perfect score – green
3	(1) gold
4	(2) silver
5	(3) bronze
6+	white

- **Plotting** panels with precipitation fields and score boxes using ggplot2

In the following two figures, it can be noticed some examples of Panelification tool with harp spatial. Figure 1 contains 1-hour cumulated precipitation for INCA data which is valid at 17 UTC from 21.06.2023 (top left), for AROME from different runs (top middle) and the solutions (control, mean and median) provided by CLAEF (top right and all 3 figures from bottom) also provided by different runs for median values. Figure 2 shows the FSS ranking (FSS skilful based on fobs=0.5) for Vorarlberg verification domain situated in the Western part of Austria.

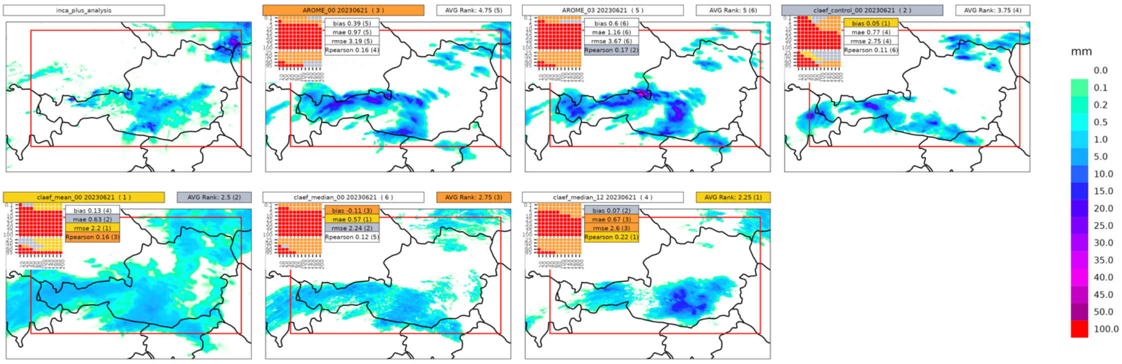


Figure 1: 1h cumulated precipitation, validity: 21.06.2023, 17 UTC, INCA on the top left, and the other plots show the forecast of AROME and CLAEF.

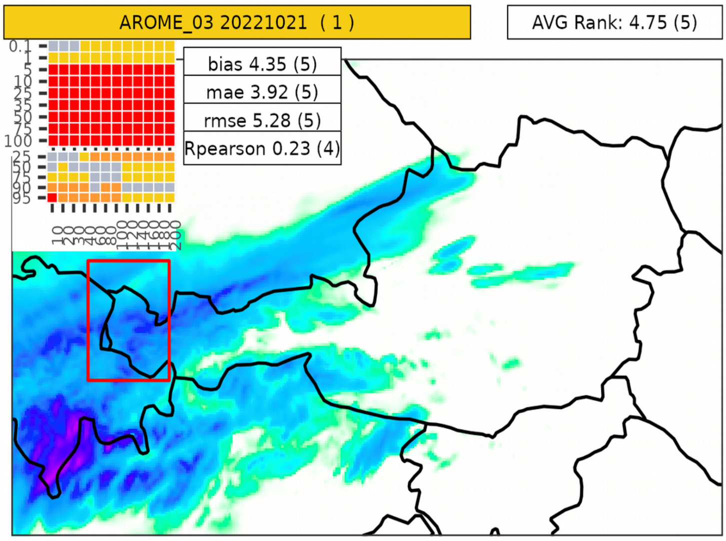


Figure 2: FSS ranking, verification domain: Vorarlberg, FSS - no NaN returned, D90 to come.

Contributer: Polly Schmederer (1pm – MQA1.2)

Using HARP for verification of AROME-EPS ***(Hungary)***

The plan is to use HARP monthly to get a detailed view of the quality of AROME-EPS forecasts. New scores, more stations, easier and more flexible data handling and plotting are available in HARP, with respect to the current self-developed EPS verification program. Therefore, the prepared scripts to verify all relevant surface and upper air parameters, for a pointwise comparison with OMSZ surface stations and available radiosonde measurements was done. Forecasts and measurements are available in GRIB and NetCDF format, respectively. For the special NetCDF files, it is used the local reading routine. The GRIB files are handled with the built-in function of HARP, wind and precipitation components are treated separately, and added after that. Hourly surface radiation was added as a new parameter, also some scores relative to average observation values were calculated. Some functions were added to plot forecasts of different initial times separately. Coverage (percentage of cases when measurement lies between highest and lowest forecasted values) score was extracted from rank histogram data. A series of plots were produced and put into PDF report using R library R Markdown. The work is ongoing, some example plots are shown in Figure 3.

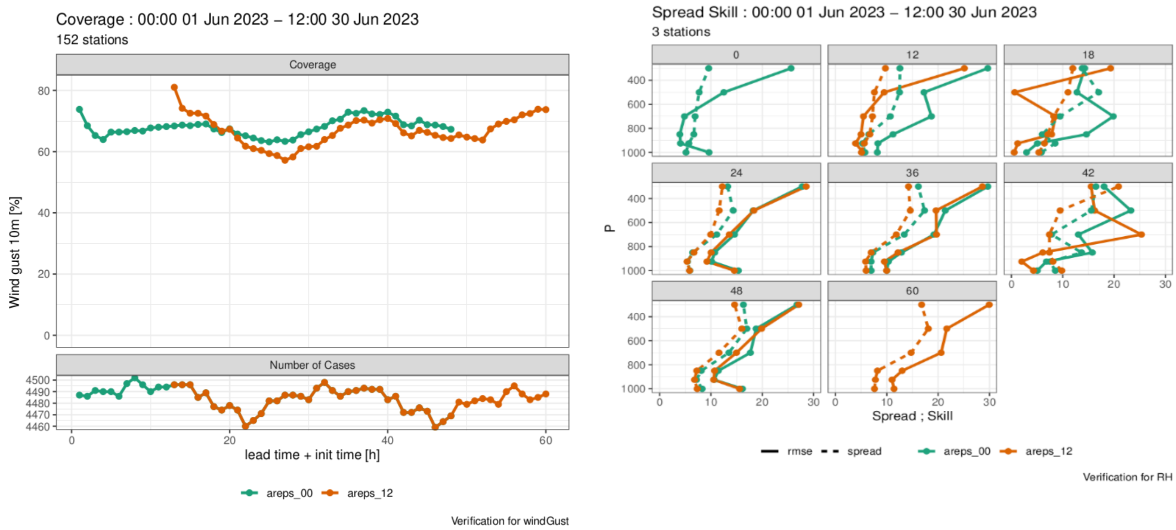


Figure 3: AROME-EPS verification scores for June 2023 for 0 UTC (green) and 12 UTC (orange) runs. Left: ensemble coverage [%] for 10 m wind gust as function of lead time (shifted by forecast initialization time). Right: RMSE of ensemble mean (solid) and ensemble spread (dashed) for relative humidity [%] as function of pressure level [hPa]. The individual panels belong to different lead times.

Contributor: Katalin Jávorné Radnóczi (0.5 pm - MQA2)

Forecast accuracy of the models **(Poland)**

In IMGW-PIB, forecasters compare forecasts from several NWP models run in an operational mode by using HARP and some other R scripts. Forecast accuracy of the models sometimes differ and can be specific to some circumstances (e.g. model A predicts the foehn effect much worse than model B). Therefore, forecasters qualitatively synthesize forecasts from different models basing mostly on their experience and knowledge. Thus, it was done a comparison of verification scores of air temperature forecasts produced by three short-ranged NWP models

- AROME 2.5 km
- ALARO 4 km

- COSMO 7km

Two of the models (AROME and ALARO) belong to the ACCORD NWP community. ALARO is based on lateral boundary conditions (LBC) coming from a global model APREGE and provides LBC to AROME. COSMO is a limited-area model from COSMO consortium and is run based on LBC from a global model ICON.

In Figure 4 it can be seen that temporal variability of bias in the considered NWP models is similar. All of the models tend to overestimate air temperature especially at nights and in the mornings during warm months (from April until September). In case of ALARO and AROME, warm bias persists also during daytime, but it is smaller than at nights. In cooler months (from October till March) cold bias occurs for all the models, which is usually little in magnitude. The greatest underestimation of temperature is forecasted by COSMO during daytime. On the contrary, the forecast error of AROME is neutral for that period.

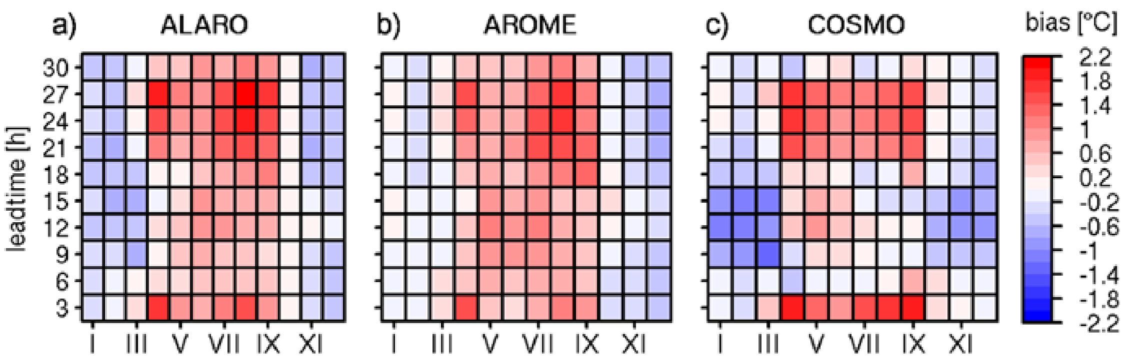


Figure 4: Temporal variability of bias in the NWP models.

Apart from seasonal and diurnal changes, variability of forecast error within a given temperature range was also examined (Figure 5). For the most commonly occurring values of air temperature (0-20°C), all the models have got relatively low RMSE (below 2°C) and the differences are quite small. For temperatures over 20°C, forecast error for COSMO gradually increases, while for ALARO – diminishes. In fact, the lowest RMSE for ALARO occurs for the highest temperatures (>30°C). For temperatures below 0°C, a distinct rise of forecast error could be observed, particularly for COSMO and AROME as their RMSE rushes to 3.85°C and 4.54°C, respectively. Interestingly, ALARO exhibits a completely opposite trend – its error diminishes as the temperature gets lower. However, due to an exceptionally small number of cases in the coldest class (most of which occurred at one station), the scores might not be relevant.

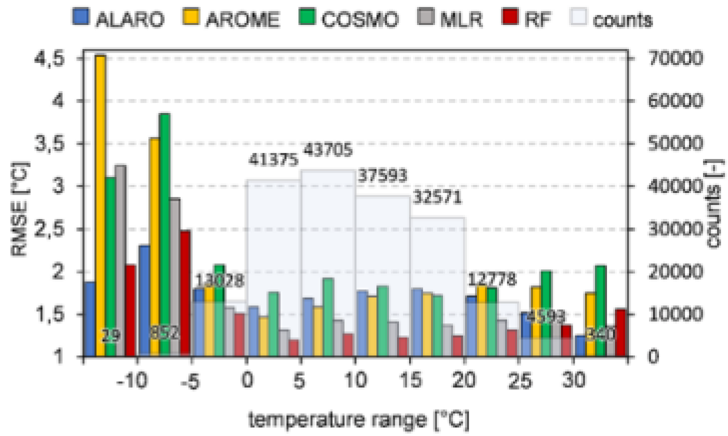


Figure 5. variability of forecast error within a given temperature range.

The last step of the analysis included investigating variability across different atmospheric circulation patterns (Figure 6). It was used a classification of atmospheric circulation types proposed by Litynski and modified by Pianko-Kuczynska with 9 cyclonic, 9 anticyclonic and 9 gradientless types (subscripts c, a and o, respectively). The largest values of RMSE for the NWP models occur for cyclonic and anticyclonic types with an undefined flow direction (Oc and Oa). Relatively high errors are noticed also in case of eastern types (especially for ALARO). It is also apparent that scores for COSMO stands out from the rest of the models for air flowing from the south (e.g. SEa, SWo, Sc). The cyclonic types of circulation from the north-east (Nec), west (Wc) and south west (SWc) have the lowest RMSE.

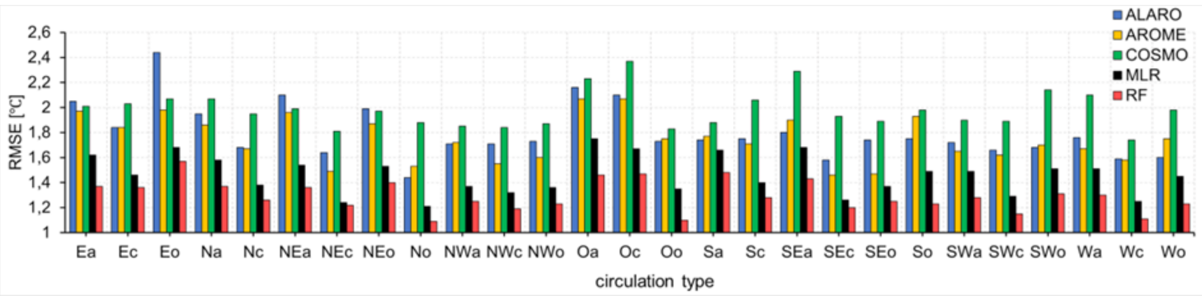


Figure 6: Variability across different atmospheric circulation patterns.

On the basis of forecasts from the NWP models, it was proposed a machine-learning-based post-processing tool to support forecasters in their decision-making process. On average, its RMSE was by 27% lower against RMSE of the most accurate NWP model (which is AROME) and by 12% - against the elementary post-processing method (which is multiple linear regression). The scores of the proposed tool are referred to in Figures 5 and 6 as *RF*.

Contributer: Gabriel Stachura (1 pm)

Contributers: Polly Schmederer (1 pm), Katalin Jávorné Radnóczi (0.5 pm – MQA2), Gabriel Stachura (1 pm)

2 Action: HARP linked to OPLACE database

In **Slovakia**

In order to read *obsoul* data, which is the specific format used by the observation preprocessing system for LACE (OPLACE), into HARP tool the work was continued by Martin Petras in a stay at CHMI in collaboration with Alena Trojáková.

For this purpose, the verification was performed by using HARP and VERAL verification systems:

- HARP version used for these experiments can be found here:
  remotes: install_github("meteorolog90/harp-develop", "develop")
- VERAL - point verification and elementary spatial analysis

Two formats of observation data:

- *obsoul* from OPLACE
- *vobs* from ECMWF observation database.

For the ease and faster way of carrying out these experiments, only two stations were chosed:

- PRAHA-RUZYNE with SID number 11518
- station PRAHA-LIBUS, with SID number 11520.

The numerical forecast data used for the validation, two deterministic models have been used:

- ALADIN-CY43 (2.3km)
- ALADIN-CY46 (2.3km) (in graph named Dakw).

In the Table 1 - left are presented the variables defined in the *obsoul* files and in right new parameters which were added in the updated version. In order for harp to read the new parameters, they must be defined in *harp_params.R*.

Table 1: Variable names and corresponding numbers (varno) in obsoul files – left, Added new variables with corresponding numbers – right.

varno	name	harp name
1	mean sea level pressure	Pmsl
39	2m temperature	T2m
58	relative humidity	RH2m
7	specific humidity 2m	q2m
41	wind speed&direction	S10m,D10m
91	cloudiness	CCtot

varno	name	harp name
79	1h precipitation accumulation	AccPcp1h
80	6h precipitation accumulation	AccPcp6h
81	minimum temperature	Tmin
82	maximum temperature	Tmax
92	snow depth	Snow

In Figure 7 can be noticed the steps for adding new parameters to Harp interfaces. The process would involve creating a list for a specified format, in this case for obsoul, where you would define:

• name: for obsoul varno
• units
• harp_name: Tmin, Tmax, T2m etc.

```
30 R/harp_params.R
@@ -124,6 +124,12 @@ harp_params <- function() {
      fa = list(
        name  = pad_string("CLSMINI.TEMPERAT", 16),
        units = "K"
+     ),
+
+     obsoul = list(
+       name = 81,
+       units = "K",
+       harp_name = "Tmin"
      )
    ),

@@ -156,6 +162,12 @@ harp_params <- function() {
      fa = list(
        name  = pad_string("CLSMAXI.TEMPERAT", 16),
        units = "K"
+     ),
+
+     obsoul = list(
+       name = 82,
+       units = "K",
+       harp_name = "Tmax"
      )
    ),
    ###
```

Figure 7: How to add new parameters into Harp interfaces.

To read and perform T2m correction from FA files, the surface geopotential name was replaced in *get_fa_param_info.R* from *SURFGEOPOTENTIEL* into *SPECSURFGEOPOTEN*.

In order to avoid the duplication obstype number (SYNOP and SHIP have the same obstype number) when the merge of the GTS and national obsouls data is done, the SHIP was removed. More details can be found in the Martin's stay report, which is available on RC-LACE web page.

Martin Petras, in his report from 2023 showed also the importance of accurately describing the station's position and he made some experiments by generating two lists in which two stations will be listed:

- the first station list gives lat and lon in short form
- second station list includes long descriptions of lat and lon

Station list with short formatting:

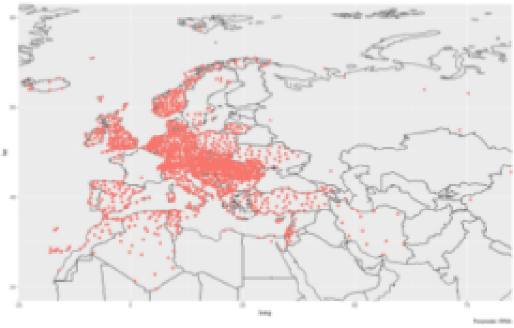
SID	lat	lon	elev	name
11518	50.1	14.26	365.0	PRAHA-RUZYNE
11520	50.0	14.45	304.0	PRAHA-LIBUS

Station list with long formatting:

SID	lat	lon	elev	name
11518	50.10028	14.25556	365.349	PRAHA-RUZYNE
11520	50.00778	14.44694	302.00	PRAHA-LIBUS

For these experiments, the following setup is run: verification period is from 01.03.2023 to 17.03.2023, with lead-time of 24 hours that were starting at 00 UTC. Verification step was 6 hours, model output file were in the FA format. Observation values was taken from obsoul and vobs files. Largest differences were found for 10m wind and smaller ones for other parameters. See more details in Martin Petras's report.

Due to the fact that are differences in geographic coverage of stations, another important step for the validation was to do a more extensive comparison of obsoul and vobs data from this point of view. The harpIO interfaces was used to check the coverage for different parameters. In the report are presented figures which contain the data coverage for different meteorological parameters: in some regions obsoul files have more data while in others vobs files have a better coverage. For 6h/1h precipitation accumulations field (Figure 8), the data availability is poor. It's not because of poor measurement coverage, but because of technical issues that needs to be resolved.



(a) 6h precipitation obsoul.

(b) 1h precipitation obosul

Figure 8: Geographical coverage of precipitation parameters over Europe.

For the comparison of HARP and VERAL scores, one station Praha-Ruzyne was selected for a one-day, 05th of May 2023 forecast starting from 00UTC. The validation was performed for several parameters: *T2m* (without height), *RH2m*, *S10m* (wind speed), *D10m* (wind direction), *CCtot* (cloudiness).

This comparison helped to identify several issues:

- cloudiness can be in fraction or oktas. For inter-comparison of Harp and VERAL the cloudiness fractions were converted into oktas scale using the "step scaling" function according to Svabik [1].
- for precipitation, the Harp can derive precipitation from other network times (for example: missing 6h precipitation at 06 UTC and 18 UTC can be derived from 12h ones reported at 06 UTC and 18 UTC minus 6h accumulations reported at 00 UTC and 12 UTC; *obsoul* already contains 6h precipitation for 00 UTC, 06 UTC, 12 UTC and 18 UTC, so this derivation part needs to be skipped). For this purpose, a new argument obs_file_format was added into read_point_obs function to accomplish this and the user needs to specify if he is using obsoul data format.
- Harp counts only four precipitation fluxes (SURFPREC.EAU.CON, SURFPREC.EAU.GEC, SURFPREC.NEI.CON, SURFPREC.NEI.GEC), while VERAL consider five of them (in addition, there is SURFPREC.GRA.GEC - the precipitation flux of new prognostic graupel.

Conclusion of Martin Petras's report are as follows:

- to test the updated read_obsoul function, which is capable of reading obsoul format. It was concluded based on the results that the updated changes have no effect on reading and verifying. It is backed up by a comparison to VERAL.
- Through testing, it was gained a better understanding of the source code, opening up new possibilities for future implementations and it started the process to add some new parameters. There was also a slight update to the FA files. It was discovered that there was a problem with reading geopotential fields from FA files. The problem has been addressed and will be solved.

- Currently, this updated version is not included in the official Harp repository. This is due to the fact that the newest versions have a number of significant changes. Our plan is to incorporate this repository into the official release after it has been released.

More details can be found in the Martin's stay report, which is available on RC-LACE web page: **https://rclace.eu/applications-and-verification**

Bibliography:

[1] Filip Svabik. Validation of harp Ecosystem: Point Verification of Surface Parameters. Tech. rep. CHMI, 2019.

Contributors: Martin Petras (1 pm), Alena Trojáková (3 weeks) – MQA1.2

3 Action: Multiple verification methods

Panelification (***Austria***)

Regarding Panelification tool some changes were made both from a technical and a scientific point of view.

1. Technical work :

- In the Framework of the DE-330 Project, the code was partially refactored to accomodate a wider range of regions and simulations.
- It was made significantly easier to add new regions and verification subdomains.
- The map generation was improved, it does not require individual border files for each country but will generate any borders and coastlines worldwide automatically.
- The panels are now checked for their aspect ratio and the paneling is adjusted accordingly.
- The score visualization was greatly improved. The scores are now shown outside of the map to not cover any part of the precipitation field. The fraction skill score ranking plots (Figure 9) now also include information on whether the model is over or underestimating the precipitation for the given window and threshold.

2. Scientific upgrades:

- Verification subdomains are now always based on a 1-km grid on a stereographic projection, and observations and model are always interpolated to this new grid, making them work the same way everywhere. This solves a previous problem, where meridian convergence and the projection of the observation grid could cause the verification domain to deviate from the desired rectangular shape.

- A second option for FSS-Visualization was added (Figure 10). This shows the deviation from the average usefull and skillful value for the given sample of models. This is a quick way to identify relative performance of models beyond just the top 3 ranks. The plot shows relative performance of the models, revealing more detailed information, like the strength of the given forecast being its good performance for strong rain while being worse than the other models for weak precipitation.
- Some work was done on the ranking, exploring alternative methods, but no decision has yet been made.

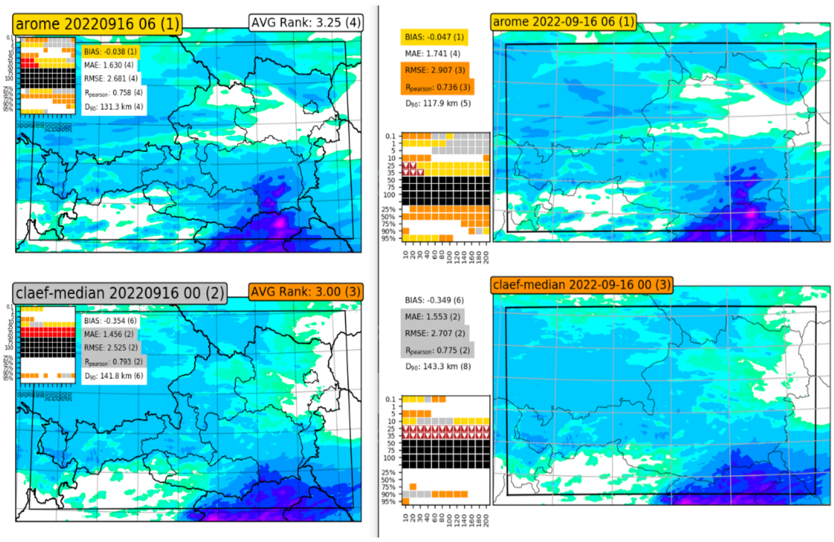


Figure 9: Comparison between the old (left) and updated (right) panels. The slight change in the shape of the verification domain is due to the updated and more general method.

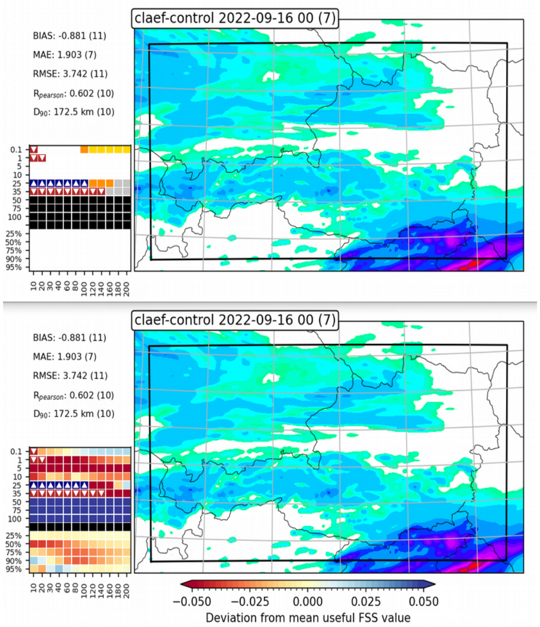


Figure 10: Showing rankings (top) and relative values (bottom) for the same panel and precipitation field.

Contributer: Phillip Scheffknecht (2 pm – MQA2)

Verification of the DE-330 53 use cases (***Czech Republic***)

Preparation of historical observation dataset for verification of the DE-330 53 use cases. It was noticed that *vobs* files currently used for verifications have gaps mainly in precipitation data. The *bufr2obs* tool was extended to extract precipitation not only from "totalPrecipitationPastXXHours" but also using "totalPrecipitationOrTotalWaterEquivalent" together with specified the accumulation period. In the Figure 11, it can be observed the illustrations of coverage of 6h precip recomputed (ec:/hirlam/oprint/OBS) on the left panel and after the new modifications on the right panel including data from Germany and Swiss. The coverage of hourly precipitation also improved considerably, see Figure 12.

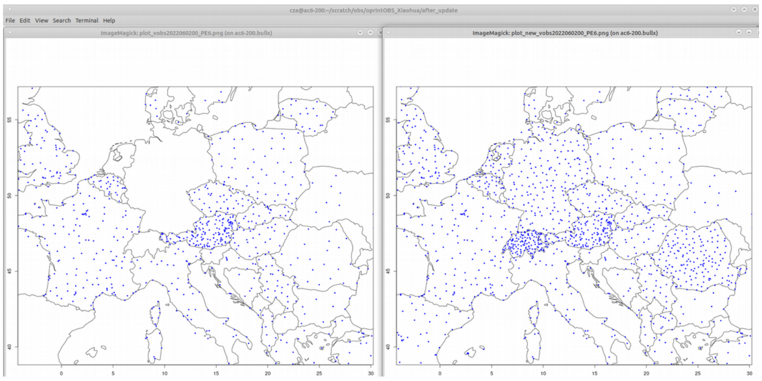


Figure 11: Coverage of 6h precip recomputed.

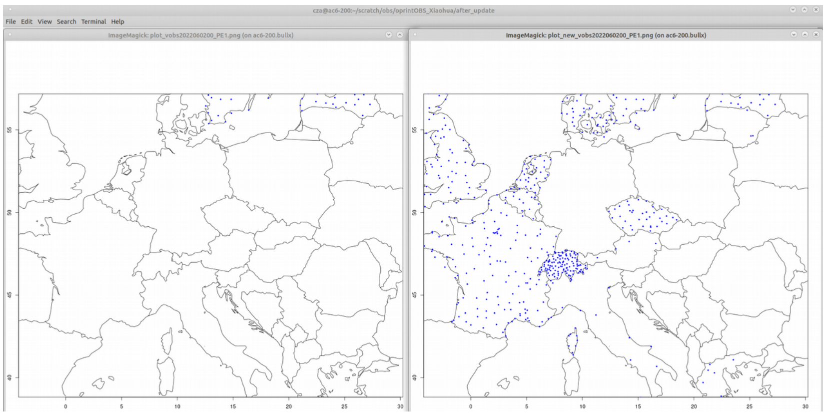


Figure 12: Coverage of 1h precip recomputed.

Contributer: Alena Trojáková (2 weeks - DE-330 53 (WP53))

"Validation of snow cover forecast by numerical weather prediction model ALADIN" by J. Sevcik - bachelor thesis ***(Czech Republic)***

The main focus of the bachelor thesis "Validation of snow cover forecast by numerical weather prediction model ALADIN" by J. Sevcik was the validation of ALADIN model snow-related variables. For that, observations of snow depth and snow water equivalent from winter season 2021/2022 have been compared to corresponding forecasts. Forecast ranges 6 and 30 hours have been selected.

For snow depth validation, only data from stations which have measured regularly every day were used, reducing the total number of available stations from 436 to 366. The loss of data caused by omitting irregularly measuring stations is the most significant for mountain stations. The data have been processed using a tool based on Python, specially developed for this work. To note that, snow depth is not a model prognostic variable and had to be estimated from forecast snow water equivalent Ws and density ρs. The comparison of the forecast and observations averaged over all stations was done. It was obtained that the the value of snow depth is underestimated by the model. In the case of the 6 hour forecast range,the total value of bias is approximately −3.4 cm and for for the 30 hour is −3.3 cm.

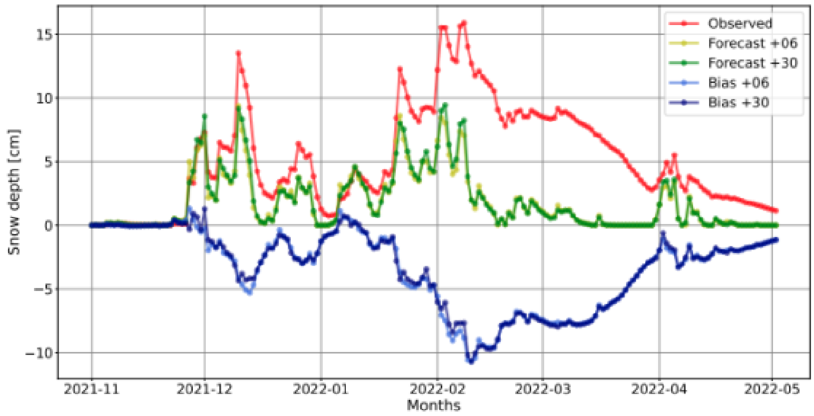


Figure 13: Daily time series of observed and forecast snow depth, averaged over all regularly measuring stations and their absolute bias, at forecast ranges 6 and 30 hours.

From Figure 13 it can be seen that the model have a tendency to underestimate the snow accumulation phase while exaggerating the melting phase. It was most pronounced during February and March. As shown in Figure 14 there are significant differences in observed snow depth among the mentioned altitude groups. The forecast error vary with altitude as well. It is shown that in the case of middle, upper and especially mountain stations, the values of model error are generally greater when the forecast is underestimated (Figure 15)

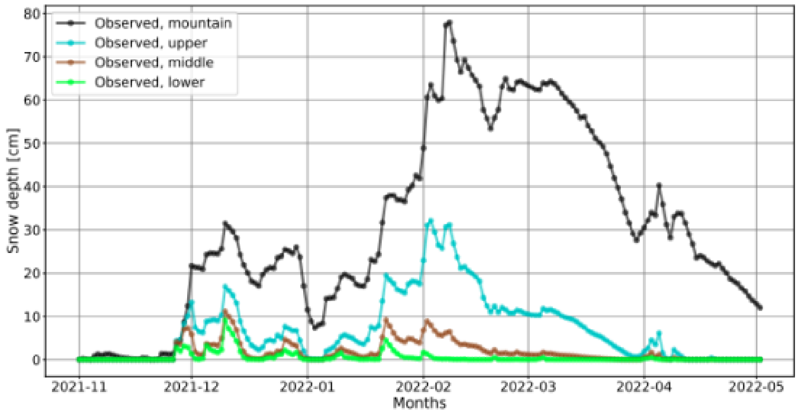


Figure 14: Daily time series of observed snow depth averaged over different altitude groups.

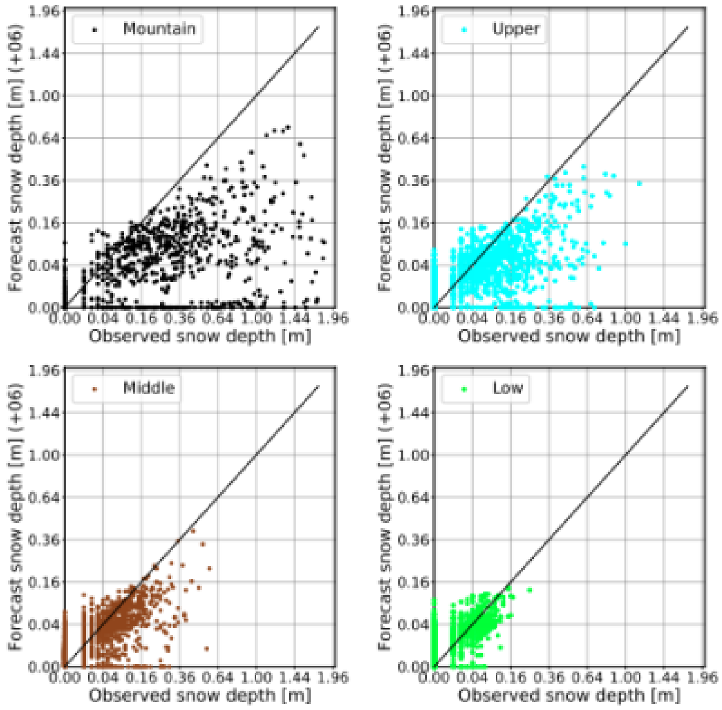


Figure 15: Scatter plots comparing forecast and observed snow depth for different altitude groups, both axes with square root scale, at 6 h forecast range.

Same results are obtained for snow water equivalent. The model underestimates forecast snow water equivalent in altitudes above 400 m a.s.l. Figure 16 show that the density forecast is quite accurate in the first half of the winter season. However, in the second half it tends to be underestimated, when snow is mostly present in mountain and upper altitudes and had settled over time.

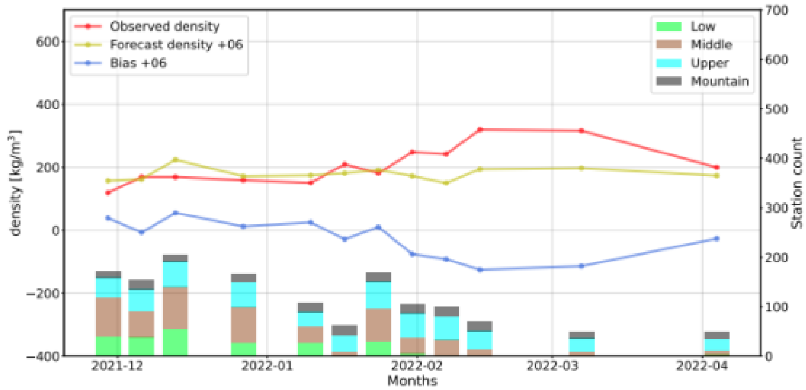


Figure 16: Time series of observed and forecast density, absolute bias, barplots with counts of available stations, at 6 h forecast range.

More details can be found in Jachym Sevcık’s bachelor thesis.

Contributers: Jachym Sevcık (1 pm), Radmila Brozkova, Jan Masek and Alena Trojáková (1 pm)

Adaptation of machine learning post-processing method on AROME forecasts (***Hungary***)

The main activity was to implement the post-processing methods developed in 2021 into the operational system. OMSZ provides forecasts from numerical weather prediction models to support partners producing renewable energy. With statistical post-processing, errors of global radiation and near-surface wind forecasts can be reduced. As part of a project, mathematician colleagues developed machine learning and EMOS (ensemble model output statistics) methods applicable to AROME and AROME-EPS ensemble forecasts in 2021.

Contributor: Dávid Tajti (1 pm)

Contributers: Phillip Scheffknecht (2 pm – MQA2), Alena Trojáková (2 weeks - DE-330 53 (WP53)), Jachym Sevcık (1 pm), Radmila Brozkova, Jan Masek and Alena Trojáková (1 pm), Dávid Tajti (1 pm)

4 Action: Post-processing of model output

In 2021, in **Hungary**, work was started on post-processing of AROME and AROME-EPS outputs. The main objective is to improve the forecasts for global radiation and 100-meter wind speed. The methodology is established and the verification is done in the framework of the regular validations.

Contributers: -

5 Action: RC-LACE Github platform

Regarding gitlub platform, it is not suitable for RC-LACE to host the platform there. A free platform, as github platform, is more convenable. A github platform dedicated for ACCORD consortium already exists. What application do you want to store on github platform?

Contributers: -

6 Action: Database of cases

Evaluation of severe weather events in summer 2023 (**Austria**)

- At Geosphere, red warning situations always trigger a procedure of evaluation of the weather situation including quality of warnings and forecasts, including model forecasts
- In summer 2023 several cases with severe weather occurred over Austria and were evaluated
- One of theses cases was the severe precipitation event that hit Slovenia but also affected parts of southern Austria

In Figure 17, top left panel shows INCA analysis of 12 hour precipitation forecast for target time 03/08 18 UTC to 04/08 06 UTC. Plots show all available AROME-Aut, CLAEF Mean/Median and IFS-HRES for target time 03/08 18 UTC to 04/08 06 UTC. All AROME-Aut and C-LAEF forecasts predicted large areas with high precipitation.

Precipitation event 03-04/08/2023 in southern Austria

- AROME-RUC forecasts with superior location of highest precipitation amounts for the first 6 hours of the event between 03/08 18 UTC to 04/08 00 UTC (Figure18)
- AROME-RUC removes overestimation of precipitation in the western part of the verification domain (black rectangle) compared to AROME-Aut/C-LAEF forecasts
- For the following 6 hours AROME-RUC didn‘t add any value to the 2.5km AROME versions.

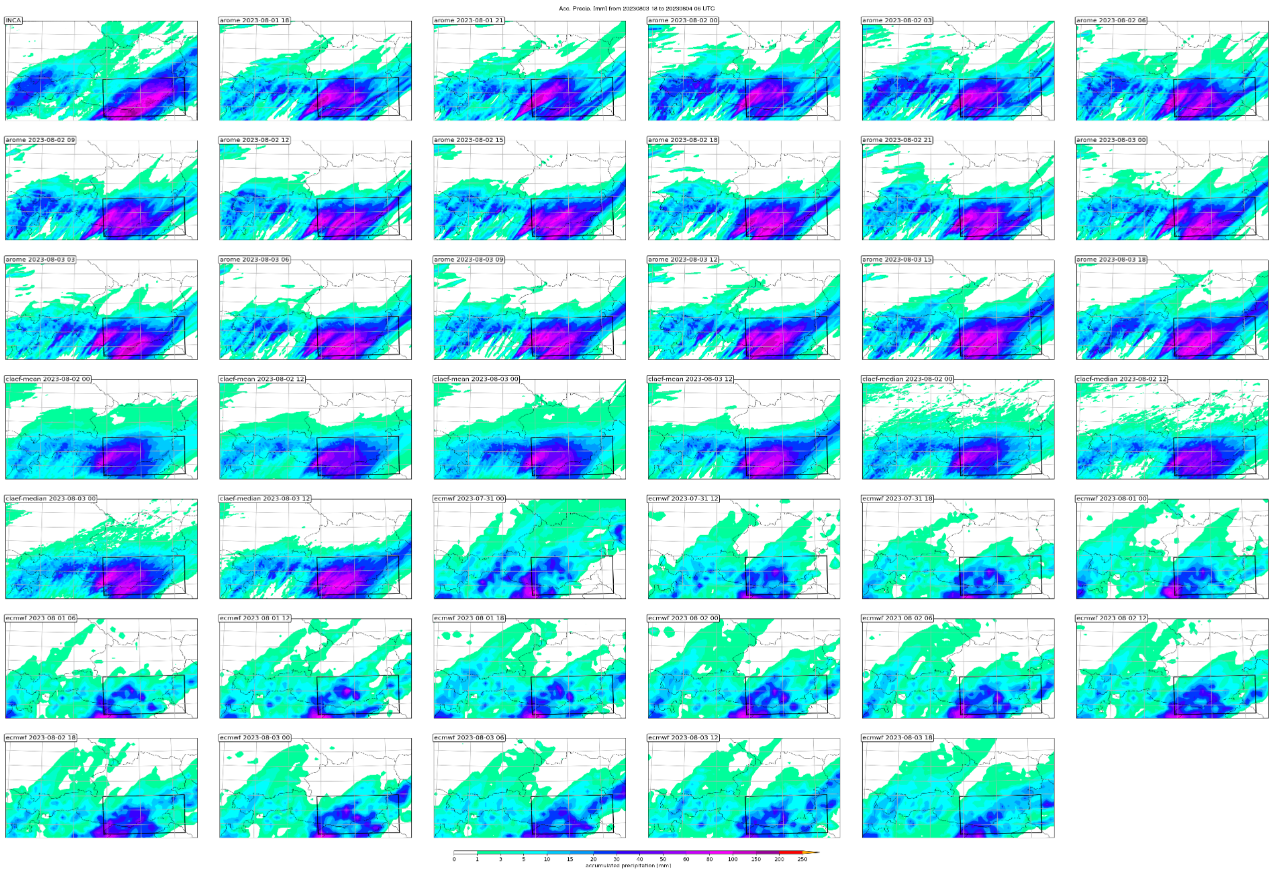


Figure 17. 12 – hour cumulated precipitation: INCA analysis and AROME and ECMWF forecasts for for target time 03/08 18 UTC to 04/08 06 UTC

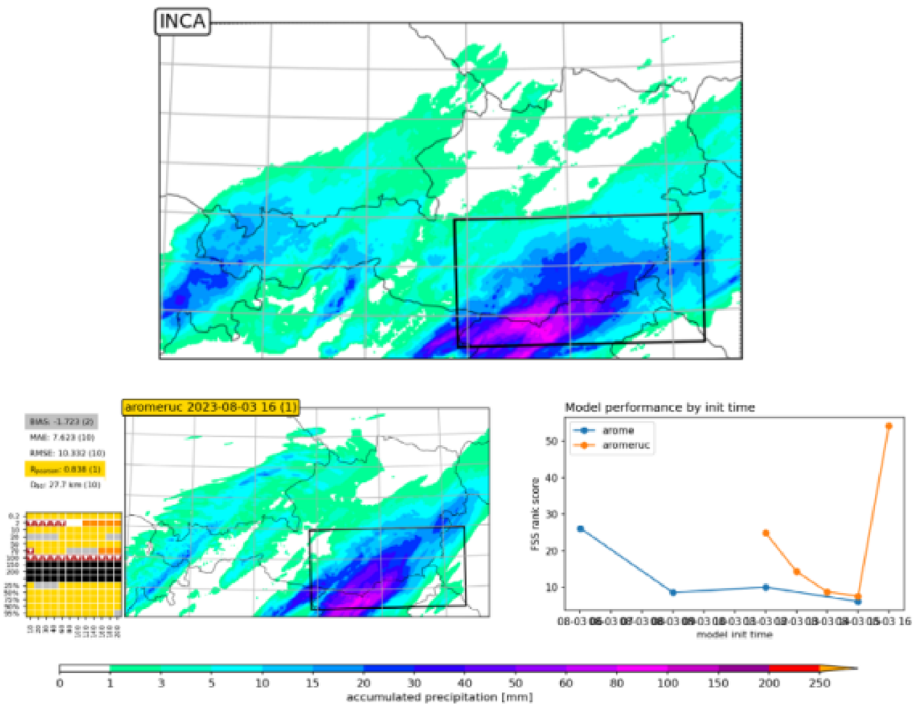


Figure 18. AROME-RUC forecasts.

Contributors: Christoph Wittmann, Clemens Wastl, Phillip Scheffknecht (3 weeks)

7 Action: Trainings

Summary of resources [PM]

Subject/Action	Resource		LACE stays	ACCORD stays (MQA?)
	planned	realized		
HARP implementation and verification for deterministic and probabilistic forecasts	3.5	2.5		
HARP linked to OPLACE database	1	1.75	1	
Multiple verification methods	3.5	5.5		
Post-processing of model output	3	-		
RC-LACE github platform	-	-		
Database of cases	1	0.75		
Trainings	-	-		
Total	12	10.5	1	