THE NWP ACTIVITIES AT CROATIAN METEOROLOGICAL AND HYDROLOGICAL SERVICE, 2025

IVAN VUJEC, JAKOV LOZUK, SUZANA PANEŽIĆ, MARTINA ĆORKOVIĆ, KRISTIAN HORVATH, MARIO HRASTINSKI, ENDI KERESTURI, ENA KOŽUL RAJČIĆ, IRIS ODAK, INES OLROM MUIĆ, ANTONIO STANEŠIĆ, MARTINA TUDOR, ANAMARIJA ZAJEC

DHMZ Croatian Meteorological and Hydrological Service

ACCORD A Consortium for COnvection-scale modelling Research and Development

Operational setup

Current operational model configurations:

- **ALADIN-HR40**: Δx=4 km; 480x432x73; CY43T2; HYD dyn.; t=150 s; ALARO-1 phy.; IC: CANARI + 3DVar (3h-cycle, ENS B); 72h fcst.; LBC: IFS-3h (6-h lagged), 4 runs per day
- **ALADIN-HR20**: Δx=2 km; 450x450x87; CY43T2; NH dyn.; DFI ini.; t=60 s; ALARO-1 phy.; 72h fcst.; IC: ALADIN-HR40; LBC: IFS 1-h (6-h lagged); 4 runs per day
- **Analog-based method**: a statistical post-processing method that identifies analogous situations in a historical (training) period using a similarity metric, and predicts future states based on corresponding past observations; predictor weight optimization and statistical correction for rare events are also applied

All-sky assimilation of IASI data

The Metop satellites are Europe's first operational meteorological satellites in polar orbit. Each Metop satellite carries the same sophisticated suite of instruments providing fine-scale global data. One such instrument is an Infrared Atmospheric Sounding Interferometer (IASI) that measures infrared energy emitted by the earth-atmosphere system in 8461 individual spectral channels with a spatial resolution of 50 km at nadir.

Conventional clear-sky data assimilation only considers data from regions without clouds, leaving the model underrepresented in cloudy areas. In preparation for the future hyperspectral infrared sounders, the assimilation of multilayer cloud-affected infrared radiances using the all-sky approach proposed by Kozo Okamoto et al. (2023) is explored. For this purpose, IASI data are used as a proxy.

The crucial part of the Okamoto approach is the observation error modeling where observation error assigned has the same size as the first guess departures standard deviation. In this sense, data is binned by the values of averaged cloud effect, defined as:

$$C_A = \frac{|B - B_{clr}| + |O - B_{clr}|}{2}$$

where B and B_clr are calculated simultaneously in RTTOV and represent model simulated brightness temperatures in cloudy and clear-sky scene for a given location. O is the observed brightness temperature. After that, linear fit to standard deviation values is performed, while defining minimum and maximum value that can be applied (where linear fit is a good approximation; **Fig 1.**).

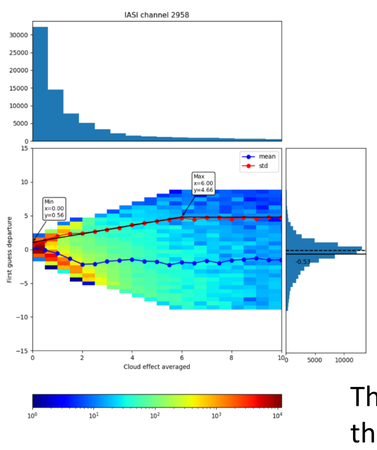

Figure 1. *2D histogram of first guess departure as a function of the averaged cloud effect; Standard deviation (red line), mean (blue line), linear fit (black line)*

The observation error is then defined as: $$O_{err} = \frac{\sigma_{cld} - \sigma_{clr}}{C_{cld} - C_{clr}}(C_A - C_{clr})$$

where σ_clr and σ_cld are minimum and maximum values of standard deviation that can be applied and C_clr and C_cld are averaged cloud effect value at which they are assigned.

Observation error modeling was technically implemented into the CY48T3. For testing purposes 3DEnVar member of the C-LAEF 1k (Convection-permitting Limited-Area Ensemble Forecasting) system was used. Four water vapor channels were selected to be assimilated in all-sky mode: 2951, 2958, 3049 and 3105, while the rest of the channels were assimilated in the clear-sky mode.

Analyzing the cases it was concluded that dynamical assignment of observation errors for all-sky channels performs as intended, and this setup can be used as the base for further development.

Temperature post-processing using Machine Learning

- HR40 NWP is used as a baseline for post-processing of temperature based on Machine Learning (ML). In the study, data from 32 meteorological stations across Croatia are used. Three years of data are used for training, while a one-year period is used for final forecast testing and evaluation.
- Three different models are analyzed:
  - Random Forest (**RF**)
  - Extreme Gradient Boosting (**XGB**)
  - Long Short-Term Memory (**LSTM**) neutral network
- Several different approaches are tested (**Fig 2.**):
  - Target variable: direct temperature forecasting " " vs **NWP error forecasting** "e" (RF, XGB, LSTM)
  - No. of models: **one model for each station** "MM" vs one model for all stations "UM" (RF, XGB)
- NWP error forecasting outperforms direct temperature forecasting for XGB and RF, while for LSTM this approach, although resulting in a slightly higher overall RMSE, performed better for extremes (not shown). For LSTM, one unified model is chosen for all stations, whereas for XGB and RF, station-specific models are developed, due to slightly lower RMSE.

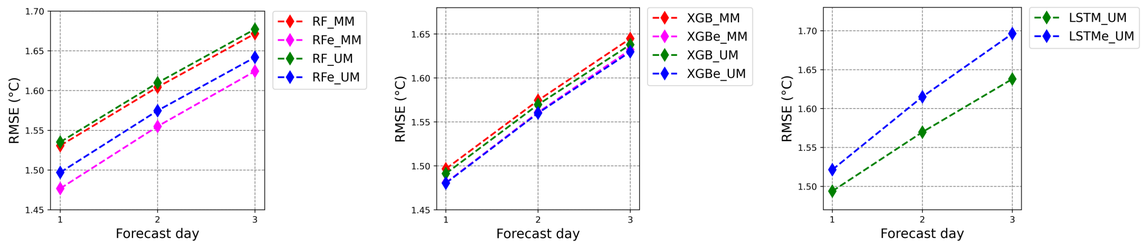

Figure 2. *RMSE for 3 forecast days ahead of different RF (left), XGB (center) and LSTM (left) models*

- All tested ML model result in **lower RMSE** for all lead time hours compared to the raw HR40 model (**Fig 3.**), and also lower RMSE than the currently operational analog-method-based post-processing.

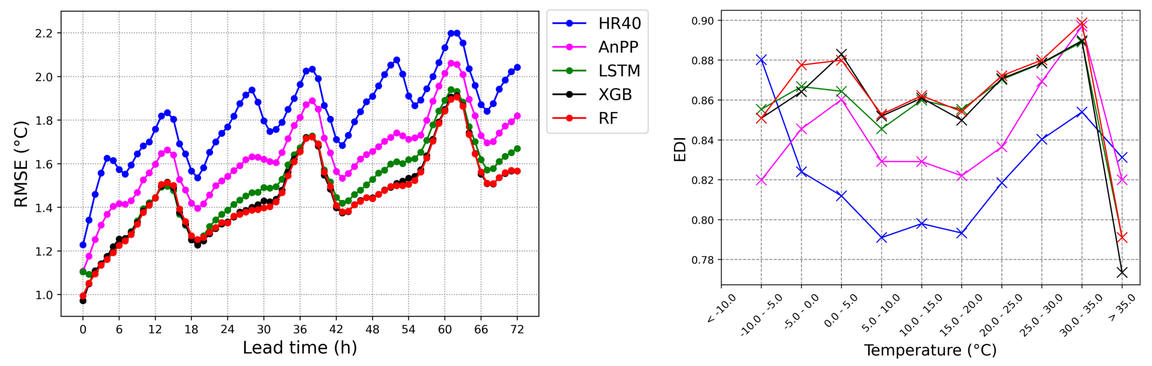

Figure 3. *RMSE for lead time hours (left) and EDI for temperature categories (right) of the raw HR40 NWP (blue line), analog-based forecast (AnPP; magenta line), LSTM (green line), XGB (black line), and RF (red line)*

- **Notable improvements in forecast performance** is spotted for the most frequent temperatures (-5-30 °C) (**Fig 3.**). For extremes, ML models still underperform compared to the NWP. Nevertheless, ML models outperform operational analog-method-based post-processing for all temperature categories except hot extremes (temperatures above 30 °C), offering a **promising perspective** for improving deterministic temperature forecasting.

LSTM visibility forecasting

- Visibility forecasting using LSTM is performed using HR40 NWP predictions and three automatic visibility measurements at one airport location (**Fig 4.**). In the study, two years of data is used for training and one year for testing and evaluation of the model.
- Visibility values, originally continuous, are divided into 7 classes based on thresholds used in aviation meteorology, due to **strong data imbalance** (**Fig 5.**).
- Classification approach: predicting visibility class rather than the visibility value itself.

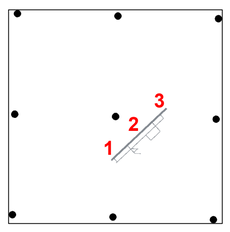

Figure 4. *Location of visibility measurements regarding the NWP grid*

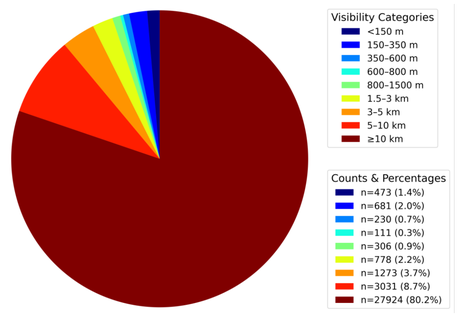

Figure 5. *Pie plot of the median of 3 visibility measurements (2020-2023)*

- LSTM generally outperforms HR40 during the first forecast day, while being comparable on the second and third forecast days in predicting low-visibility (visibility below 2 km) occurrence in 12-hour windows (**Fig 6.**). As **Fig 7.** shows, the LSTM forecast results in a higher number of Hits and fewer Misses, but at the cost of more False Alarms and fewer Correct Negatives than the raw HR40.

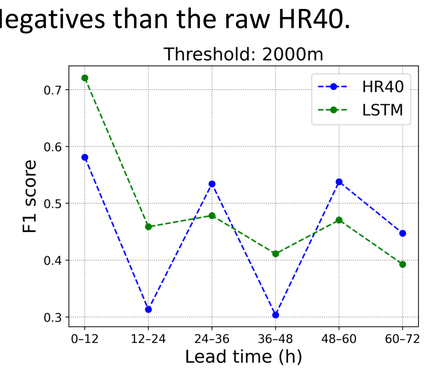

Figure 6. *F1 score of low-visibility occurrence for six 12-hour-long forecast windows for HR40 (blue) and LSTM (green)*

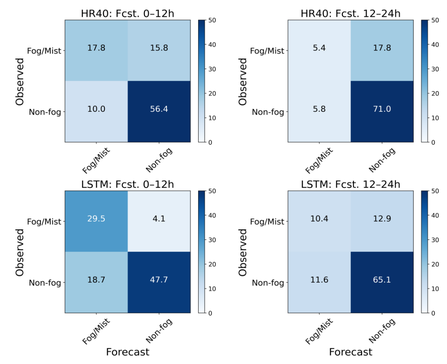

Figure 7. *Confusion matrices for HR40 (upper row) and LSTM (lower row) in the first two 12-hour forecast windows*

- LSTM shows to be a **promising tool for predicting low-visibility conditions**, especially in the **first 24 hours**. It also provides the possibility of probabilistic output, valuable information for many end-users who rely on accurate visibility forecasts.

Verification of operational ALADIN forecasts

- Two articles regarding the verification of operational ALADIN forecasts were submitted to HMČ (Hrvatski Meteorološki Časopis- Croatian Meteorological Journal). The results of post processed analog-based forecast HRAN (with predictor weight optimization and correction for more extreme events included) is compared to the HR40 (for temperature variable) and HR20 (for wind speed variable) forecasts for the year of 2024.
- Results confirm that HRAN consistently reduces overall forecast errors, primarily by lowering dispersion errors, in comparison to HR20 and HR40. (**Fig 8.**) Spatial analyses confirm that HRAN improves forecast performance across all station groups, with especially notable benefits for continental sites (for wind speed) and mountainous region (for temperature).

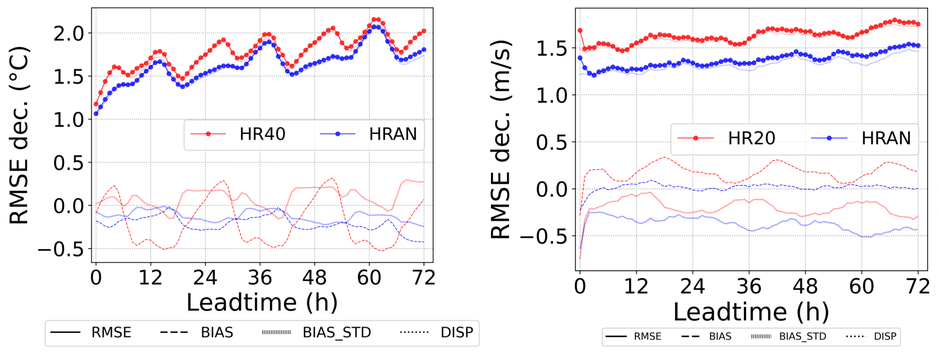

Figure 8. *RMSE decomposition for different leadtimes for temperature (left) and wind speed (right) forecasts. Data refer to forecasts initialized during 2024 at locations across Croatia.*

Categorical analyses confirm that HRAN generally improves forecast performance across most of the categories. For wind speed, the improvement of HRAN in comparison to the HR20 is evident for all categories, especially for high wind speed events. For the temperature variable, for very low or very high temperature categories, HR40 still shows superior results to HRAN.