Environment. Technology. Resources. Rezekne, Latvia
Proceedings of the 16th International Scientific and Practical Conference. 2025. Volume I, 509-514

Evaluating the Performance of EFAS Hydrological Predictions in Latvian River basins: A Comparison with Observational Data

Andris Skromulis
Rezekne Academy of Riga Technical University
Riga, Latvia
andris.skromulis@ra.lv

Juris Breidaks
Forecast and Climate Department
Climate and Numerical Modelling Division
SLLC "Latvian Environment, Geology and Meteorology Centre"
Riga, Latvia
Juris.Breidaks@lvgmc.lv

Svetlana Aņiskeviča
Forecast and Climate Department
SLLC "Latvian Environment, Geology and Meteorology Centre"
Riga, Latvia
svetlana.aniskevica@lvgmc.lv

Līga Klints
Forecast and Climate Department
SLLC "Latvian Environment, Geology and Meteorology Centre"
Riga, Latvia
liga.klints@lvgmc.lv

Darja Hudjakova
Forecast and Climate Department
SLLC "Latvian Environment, Geology and Meteorology Centre"
Riga, Latvia
darja.hudjakova@lvgmc.lv

***Abstract* — This study evaluates the performance of the European Flood Awareness System (EFAS) [1] in predicting hydrological variables by comparing EFAS reforecast data with observational data from the Latvian Environment, Geology and Meteorology Centre (LVGMC). Using the open-source LISFLOOD hydrological model [2], the study examines the accuracy of ECMWF-driven predictions of river discharge and water levels across Latvia's diverse river basins. The study employs a variety of interpolation techniques, including linear interpolation and nearest neighbour interpolation, to extract grid data from the Copernicus Early Warning Data Store (EWDS) [3] dataset at hydrological station points. To assess prediction accuracy, a range of statistical and error metrics, including Mean Error (ME) [4], [5], Root Mean Squared Error (RMSE) [5] - [7], Nash-Sutcliffe Efficiency (NSE) [5], [8]-[12] and Kling-Gupta Efficiency (KGE) [5], [12], [13], are utilized. The analysis highlights the effectiveness of EFAS in different seasonal and hydrometeorological conditions, identifying both strengths and limitations in the model's performance. Furthermore, the study explores potential calibration approaches to including regional forecasting capabilities, particularly in light of climate change impacts on low-flow and drought period predictions. This research provides valuable insights into the application of continental-scale hydrological models at the regional level, offering recommendations for improving the accuracy of flood forecasting systems.**

Keywords* — *EFAS, hydrology, REFORECAST, ECMWF, verification

I. INTRODUCTION

Predictive analysis – in other words, combining observations with exemplary information to reconstruct weather and climate – plays a major role in numerical weather forecasting. This study will include this dataset, which provides the grid simulated hydrological time. The data consistently reflect the most important hydrological variables in the field of the European Flood Awareness System (EFAS) [1].

The data reflect the most important hydrological variables in the European flood awareness system (EFAS). This dataset has been created using the open source LISFLOOD hydrological model [2] at 1X1 arcade resolution (~ 1.5 km EFAS latitude) with meteorological overforecasts of the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble. Reforecasts are forecasts that run on past dates and are typically used to

Online ISSN 2256-070X
https://doi.org/10.17770/etr2025vol1.8705

evaluate the proficiency of the forecast system or to develop tools to correct forecast statistical errors.

Hydrological data is critical for building flood prediction models. Hydrological modelling can help predict the extent and duration of flooding, which is essential for alerting the population of affected areas in a timely manner. Hydrological data are also essential for flood risk management and long-term mitigation strategies. In the context of climate change, the low-flow period and drought forecast are also significantly increasing. Each region has a unique hydrological profile made up of topography, climate, land use and other factors. Reforecasts of river discharges are proposed twice a week with lead time of up to 46 days, with 6-hour periods of 20 years [3].

Recent research has highlighted challenges in converting continental-scale hydrological predictions into regional and local applications. Arnal et al. [14] demonstrated that the skill of seasonal hydrological forecasts varies significantly across regions of Europe, significantly limiting areas with complex terrain. So Wanders et al. [15] found that reduction methods can significantly improve the accuracy of local projections when applied to continental models. Pappenberger et al. [16] further underlined the importance of a comprehensive verification system for the assessment of hydrological forecasting systems at several spatial scales. Verification of EFAS forecasts against observation data from national hydrological monitoring networks, such as the one maintained by State Limited Liability Company "Latvian Environment, Geology and Meteorology Centre" (LVGMC), provides important insights into the performance of models at different spatial scales and hydrometeorological conditions.

In the field of hydrology accurate methods of data retrieval and interpolation are essential for reliable analysis and forecasting.

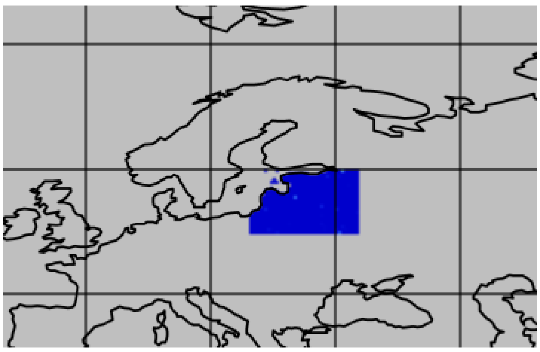

Fig. 1. Area of the EFAS reforecasts of river discharges.

If your figure has two parts, incorporate the labels "(a)" and "(b)" in the figure. At the same time, do not incorporate captions in the figures. Do not put captions in "text boxes" linked to the figures. Do not put borders around the outside of your figures. Use the abbreviation "Fig. 1" even at the beginning of a sentence.

The purpose of this article is to address an important issue in hydrological data analysis by comparing EFAS predictions with observational data from LVGMC.

This research aims to quantify the accuracy, reliability, and systematic biases of ECMWF-driven LISFLOOD predictions when compared to ground-truth measurements from Latvia's comprehensive monitoring network. By examining performance across different seasons and river basins, this study will identify specific conditions under which model performance excels or deteriorates. Furthermore, the analysis will explore potential calibration approaches to improve regional forecasting capabilities based on identified discrepancies between predicted and observed data.

II. MATERIALS AND METHODS

A. Data Extraction

The study involves 7 years (2013-2019) of EWDS grid data (ECMWF reanalysis dataset) mining at geographical data points of hydrological stations. Since the choice of interpolation method significantly impacts data accuracy, this study comprehensively evaluates the effectiveness of different interpolation techniques.

For data retrieval, we utilized the griddata function from the Python package xarray [17], [18] to extract point data by applying three distinct grid data interpolation methods:

- Linear interpolation: This method, widely used in scientific computing, estimates values at unknown points by fitting a straight line between known data points. Implementation was carried out using SciPy [19], a fundamental Python library for numerical computing.
- Nearest neighbour interpolation: A computationally efficient method that assigns the value of the closest grid point to the target location. This approach prioritizes processing speed over smoothness of results, making it particularly useful for large datasets where computational resources are limited.

B. Quality Assessment

A comprehensive set of errors and statistical indicators is used to assess the quality of the data obtained, each providing insights into different aspects of interpolation performance.

C. Basic Error Metrics

- Mean Error (ME) [4],[5]:

$$ME = \frac{1}{n}\sum_{i=1}^{n}(F_i - O_i) \qquad (1)$$

The ME measures the average difference between forecast data (F) and observed data (O). A value closer to zero indicates better performance, with positive values indicating overestimation and negative values indicating underestimation.

- Root Mean Squared Error (RMSE) [5]- [7]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(F_i - O_i)^2} \quad (2)$$

- Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{F_i - O_i}{O_i}\right| \quad (3)$$

D. Correlation Metrics

- Spearman's Rank Correlation (R_Spearman) [5], [20]:

$$R_{Spearman} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(R(O_i)-\overline{R(O_i)}\right)\left(R(F_i)-\overline{R(F_i)}\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(R(O_i)-\overline{R(O_i)}\right)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(R(F_i)-\overline{R(F_i)}\right)^2}} \quad (4)$$

This non-parametric measure assesses the strength and direction of monotonic relationships between forecast and observed data, making it robust against outliers and non-linear relationships.

- Anomaly Correlation Coefficient (ACC) [5], [21]-[23] :

$$ACC = \frac{1}{n}\frac{\sum_{i=1}^{n}(F_i - \bar{F})(O_i - \bar{O})}{\sigma_O \sigma_F} \quad (5)$$

ACC measures the correlation between the variation patterns of forecasted and observed data. A value closer to 1 indicates that the forecast accurately captures the observed data's anomaly patterns.

E. Advanced Efficiency Metrics

- Watterson's M (Watt_M) [5], [8]

$$Watt\_M = \left(\frac{2}{\pi}\right) sin^{-1}\left(1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(F_i - O_i)^2}{\sigma_F^2 + \sigma_O^2 + (F - \bar{O})^2}\right) \quad (6)$$

This metric provides a normalized measure of agreement between forecasts and observations.

- Nash-Sutcliffe Efficiency (NSE) [5], [9]-[12]:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(F_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \quad (7)$$

This metric provides a normalized measure of agreement between forecasts and observations.

- Modified Nash-Sutcliffe Efficiency (NSE_{mod}) [5], [8]:

$$NSE_{mod} = 1 - \frac{\sum_{i=1}^{n}|F_i - O_i|^j}{\sum_{i=1}^{n}|O_i - \bar{O}|^j} \quad (8)$$

This modification to the NSE formula allows for adjusting sensitivity to outliers through the parameter j, with j=1 reducing sensitivity and higher j values increasing it.

- Relative Nash-Sutcliffe Efficiency (NSE_{rel}) [5], [8]:

$$NSE_{rel} = 1 - \frac{\sum_{i=1}^{n}\left|\frac{F_i - O_i}{O_i}\right|^2}{\sum_{i=1}^{n}\left|\frac{O_i - \bar{O}}{O}\right|^2} \quad (9)$$

This version of NSE is particularly useful for evaluating relative differences, reducing the influence of magnitudes.

F. Kling-Gupta Efficiency Metrics

- Kling-Gupta Efficiency Metrics (KGE_{2009}) [5], [12]:

$$KGE_{2009} = 1 - ED_9 \quad (10)$$

where

$$ED_9 = \sqrt{\left(s[1]\cdot(r-1)\right)^2 + \left(s[2]\cdot(\alpha-1)\right)^2 + \left(s[3]\cdot(\beta-1)\right)^2}, \quad (11)$$

$$r = \frac{\sum_{i=1}^{n}(O_i - \bar{O})(F_i - \bar{F})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2}\sqrt{\sum_{i=1}^{n}(F_i - \bar{F})^2}}, \quad (12)$$

$$\beta = \frac{\bar{O}}{\bar{F}}, \quad (13)$$

$$\alpha = \frac{\sigma_F}{\sigma_O}, \quad (14)$$

where s represents the scaling factors to be used for re-scaling the Pearson product-moment correlation coefficient (r), α, and β, respectively, σ_O and σ_F are the standard deviations of O and F respectively.

Gupta et al. [11] created this metric to demonstrate the relative importance of the three components of the NSE, which are correlation, bias and variability.

- Kling-Gupta Efficiency 2012 (KGE_{2012}) [5], [13]:

$$KGE_{2012} = 1 - ED_{12} \quad (15)$$

where

$$ED_{12} = \sqrt{\left(s[1]\cdot(r-1)\right)^2 + \left(s[2]\cdot(\gamma-1)\right)^2 + \left(s[3]\cdot(\beta-1)\right)^2} \quad (16)$$

$$\gamma = \frac{\sum_{i=1}^{n} F_i \sqrt{\sum_{i=1}^{n}(F_i - \bar{F})^2}}{\sum_{i=1}^{n} O_i \sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2}} \quad (17)$$

This 2012 modification of KGE reduces the cross-correlation between bias and variability ratios, providing to more independent assessment of these components.

G. Additional Metrics

- Lengendre-Marsden Index (LM Index) [5], [24]:

$$LM_{index} = 1 - \frac{\sum_{i=1}^{n}|F_i - O_i|}{\sum_{i=1}^{n}|O_i - \bar{O}|} \quad (18)$$

This index is less sensitive to extreme values than NSE, providing a more robust assessment for data with outliers.

III. Results and discussion

For data retrieval, the nearest neighbour interpolation method gave better results than linear interpolation. Turned out that difference between upstream basin areas between in first step selected retrieved data and existing database information differs by 0.00027 to 13.48%. It was accepted that basin area difference can be no larger than 0.2%. This area difference can be linked to specific subbasin division methods.

TABLE 1 STATISTICAL INDICATORS FOR ANALYSED NEAREST NEIGHBOUR INTERPOLATION METHOD DATA

Season	Statistical Indicators											
	ME	*RMSE*	*MAPE*	*R_Spearman*	*ACC*	*Watt_M*	*NSE*	*NSE mod*	*NSE rel*	*KGE (2009)*	*KGE (2012)*	*LM Index*
DJF min	-80.80	0.35	0.32	0.12	-0.07	-0.03	-0.63	-0.39	-19.54	-0.27	-0.35	-0.39
DJF max	15.16	134.46	0.94	0.80	0.79	0.53	0.57	0.42	0.83	0.67	0.72	0.42
MAM min	-137.70	1.10	0.22	0.35	0.40	0.12	-1.43	-0.46	-1.28	-0.02	-0.54	-0.46
MAM max	17.75	195.98	0.57	0.92	0.93	0.66	0.76	0.44	0.83	0.78	0.79	0.44
JJA min	-36.63	0.35	0.41	0.06	0.14	0.05	-36.41	-4.52	-90.65	-4.11	-2.46	-4.52
JJA max	32.81	118.59	1.20	0.78	0.89	0.58	0.70	0.47	0.62	0.67	0.68	0.47
SON min	-0.45	0.38	0.42	0.25	-0.05	-0.02	-17.56	-2.66	-86.10	-3.94	-2.82	-2.66
SON max	39.21	78.38	1.38	0.89	0.90	0.65	0.73	0.41	0.80	0.82	0.82	0.41

The analysis focuses on the data comparison using statistical indicators. The data were analyzed for different seasons, as they are commonly used in hydro-climatological studies: winter (DJF), spring (MAM), summer (JJA), and autumn (SON). Runoff variation across seasons can be significant and driven by different factors. Seasonal data analysis can provide insights into periods when the model performs most effectively.

TABLE 2 STATISTICAL INDICATORS FOR ANALYSED LINEAR INTERPOLATION METHOD DATA

Season	Statistical Indicators											
	ME	*RMSE*	*MAPE*	*R_Spearman*	*ACC*	*Watt_M*	*NSE*	*NSE mod*	*NSE rel*	*KGE (2009)*	*KGE (2012)*	*LM Index*
DJF min	-91.36	0.34	0.31	0.12	-0.07	-0.03	-0.57	-0.36	-9.44	-0.26	-0.33	-0.36
DJF max	2.79	139.54	0.93	0.80	0.79	0.56	0.60	0.42	0.83	0.71	0.75	0.42
MAM min	-170.57	1.11	0.23	0.35	0.40	0.11	-1.42	-0.51	-1.28	-0.04	-0.55	-0.51
MAM max	0.14	234.31	0.60	0.92	0.93	0.66	0.76	0.44	0.89	0.77	0.77	0.44
JJA min	-58.83	0.31	0.41	0.06	0.14	0.08	-11.03	-1.92	-30.82	-1.68	-1.32	-1.92
JJA max	24.52	136.82	1.16	0.78	0.89	0.58	0.65	0.42	0.71	0.71	0.67	0.42
SON min	-1.23	0.36	0.41	0.25	-0.05	-0.02	-5.19	-1.19	-31.25	-1.55	-1.02	-1.19
SON max	28.29	46.73	1.37	0.89	0.90	0.66	0.76	0.48	0.86	0.83	0.83	0.48

The statistical analysis results for both the linear and nearest neighbor interpolation methods, based on the retrieved data compared to the observed data, were similar. Table 1 summarizes the seasonal minimum and maximum values for the nearest neighbor interpolation dataset, while Table 2 summarizes the statistical indicator values for the linear interpolation dataset. The highest differences in RMSE were observed in spring, which can be explained by the fact that the highest values and the largest value amplitudes are typically observed or modeled during this period. Also, spring season can be characterized as period with the highest correlation values for different correlation metrics - Spearman's Rank Correlation (R_Spearman) up to 0.93, Kling-Gupta Efficiency Metrics 0.77-0.78, NSE 0.44-0.89, lowest values show NSEmod, which reacts to outliers in data.

When considering the best correlation performance, the lowest correlation for both retrieved datasets compared to the observed data occurred in the two seasons most exposed to low-flow periods—winter and summer. This can be explained by the fact that the model is more suitable for high-flow periods than for low-flow periods. The results in this study were obtained from the entire set of stations. Autumn season data correlations showed good performance; many indicators were higher than 0.8 for both data sets. Autumn showed the best performance for RMSE values.

When exploring the results for different hydrological stations separately, it becomes evident that the dataset includes stations with very good performance as well as stations with poor performance. This variation occurs both from year to year and between different seasons.

Data quality is highly related to river basin properties – catchment size, soils, land-use changes, deforestation, urbanization and many others. All these factors give impact, so data analysis could be done more detailed – by sorting hydrological stations into respectively hydrological regions or grouped by different classes to to achieve more comprehensive results.

This study showed that most of stations gave satisfying correlation results for all seasons and in lack of available observation values modelled values can be the valuable substitute. However, given that the correlation results for some stations were not acceptable, one must remain cautious when using this data, as data inconsistencies may exist.

IV. CONCLUSION

This study offers examination of modelled data comparison to observed data by different statistical indices. Key findings indicate that best performance is for high flow periods, but less correlation is seen in low-flow periods – winter and summer. In addition, the study highlights differences between different station performance.

According to the interpretation by authors of this paper for many stations correlation between analysed data sets is strong enough to be practically applicable.

REFERENCES

[1] C.Mazzetti, et al., EFAS v5.0 hydrological reanalysis, European Commission, Ispra, 2023, JRC134686.

[2] J.M. van Der Knijff, J. Younis and A.P.J. De Roo, LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation. International Journal of Geographical Information Science, 2008, 24(2), pp.189–212. https://doi.org/10.1080/13658810802549154

[3] Reforecasts of river discharge and related data by the European Flood Awareness System, https://ewds.climate.copernicus.eu/datasets/efas-reforecast?tab=overview

[4] R.A. Fisher, A Mathematical Examination of the Methods of Determining the Accuracy of an Observation by the Mean Error, and by the Mean Square Error. Monthly Notices of the Royal Astronomical Society, 1920, 80, pp. 758 - 770.

[5] W.Roberts, G. Williams, E., Jackson, E. Nelson and D. Ames, Hydrostats: A Python Package for Characterizing Errors between Observed and Predicted Time Series. Hydrology, 2018, 5(4) 66, doi:10.3390/hydrology5040066

[6] C.J. Willmott and K. Matsuura, K., Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 2005, 30(1), pp. 79-82.

[7] R.J. Hyndman and A.B. Koehler, Another look at measures of forecast accuracy. International Journal of Forecasting, 2006, 22(4), pp. 679-688.

[8] P.Krause, D. Boyle and F. Bäse, Comparison of different efficiency criteria for hydrological model assessment. Advances in geosciences, 2005, 5, pp. 89-97.

[9] R.H. McCuen, Z. Knightand A.G. Cutter, Evaluation of the Nash-Sutcliffe Efficiency Index. Journal of Hydraulic Engineering, 2006.

[10] J.E. Nash and J.V. Sutcliffe, River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology, 1970, pp. 282-290.

[11] C.J. Willmott, S.M. Robeson and K. Matsuura, A refined index of model performance. International Journal of Climatology, 2012, 32(13) pp. 2088-2094.

[12] H. V. Gupta, H. Kling, K.K. Yilmaz and G.F. Martinez, Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 2009, 377(1-2), pp. 80-91.

[13] H. Kling, M. Fuchs and M. Paulin, Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. Journal of Hydrology, 2012, 424, pp.264-277.

[14] L. Arnal, H.L.Cloke, E. Stephens, F.Wetterhall, C. Prudhomme, J. Neumann, B. Krzeminski, and F. Pappenberger, Skilful seasonal forecasts of streamflow over Europe? Hydrology and Earth System Sciences, 2018, 22, pp. 2057-2072.

[15] N. Wanders, S. Thober, R. Kumar, M. Pan, J. Sheffield, L. Samaniego and E.F. Wood, Development and evaluation of a pan-European multimodel seasonal hydrological forecasting system. Journal of Hydrometeorology, 2019, 20, pp. 99-115.

[16] F. Pappenberger, H.L. Cloke, D. J. Parker, F. Wetterhall, D.S. Richardson and J. Thielen, The monetary benefit of early flood warnings in Europe. Environmental Science & Policy, 2015, 51, pp. 278-291.

[17] S. Hoyer and J. Hamman. Xarray: N-D labeled arrays and datasets in Python. Journal of Open Research Software, 2017. URL: doi:10.5334/jors.148.

[18] S. Hoyer, C. Fitzgerald, J. Hamman, and others. Xarray: v2022.3.0. May 2022. doi:10.5281/zenodo.59499.

[19] P. Virtanen, R. Gommers, T.E. Oliphant et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods, 2020, 17, pp. 261–272. https://doi.org/10.1038/s41592-019-0686-2

[20] C. Spearman, "The proof and measurement of association between two things". American Journal of Psychology, 1904, 15: 72–101. doi:10.2307/1412159

[21] R.H. Langland, and R. N. Maue. "Recent Northern Hemisphere Mid-Latitude Medium-Range Deterministic Forecast Skill." Tellus A: Dynamic Meteorology and Oceanography 64, no. 1, 2012, pp.17531.

[22] K. Miyakoda, G. D. Hembree, R. F. Strickler, and I. Shulman. "Cumulative Results of Extended Forecast Experiments I. Model Performance for Winter Cases." Monthly Weather Review 100, no. 12, 1972, pp. 836–55.

[23] A. H.Murphy and E. S. Epstein. "Skill Scores and Correlation Coefficients in Model Verification." Monthly Weather Review 117, no. 3, 1989, pp. 572–82.

[24] D.R. Legates and Jr, G.J. McCabe, Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. Water Resources Research, 1999, 35(1) pp.233-241.

[25] E.L. Lehmann and G. Casella, Springer Texts in Statistics. Springer-Verlag, New York, 1998.