ICFM 10
Adapting to Global Change:
Innovative Approaches to Flood Management and Resilience
May 20-22, 2026 | Western University, London, Ontario, Canada

Uncertainties in ML-based Daily Reservoir Inflow Forecasting: A Case Study on Drina-Lim Hydropower Plants

Luka Vinokić[1], Milan Dotlić[1], Ana Samac[1], Vanja Švenda[1], Veljko Prodanović[1,2], Milan Stojković[1,3]
The Institute for Artificial Intelligence R&D, Fruškogorska 1, 21000 Novi Sad, Serbia[1]
E-mail: luka.vinokic@ivi.ac.rs
E-mail: milan.dotlic@ivi.ac.rs
E-mail: ana.samac@ivi.ac.rs
E-mail: vanja.svenda@ivi.ac.rs
E-mail: veljko.prodanovic@ivi.ac.rs
E-mail: milan.stojkovic@ivi.ac.rs

University of South Wales (UNSW), School of Civil and Environmental Engineering, Sydney, NSW 2052, Australia [2]
E-mail: v.prodanovic@unsw.edu.au

Faculty of Civil Engineering, University of Novi Sad, Kozaračka 2a, 24000 Subotica, Serbia[3]
E-mail: milan.stojkovic@gf.uns.ac.rs

ABSTRACT

Accurate streamflow forecasting is essential for hydropower optimization and flood risk reduction. This study develops an operational, data-centric forecasting framework integrating three machine learning architectures—LSTM, TCN, and TKAN—trained on long-term hydrometeorological data and driven by numerical weather predictions from IFS, GFS and OM. Two bias correction methods, delta-scaling and quantile mapping, are applied to address systematic NWP biases. An ensemble of 27 model-forcing-bias correction combinations is used for each station to quantify predictive uncertainty, while factorial ANOVA decomposes contributions from ML model choice, NWP source, and BCM (with additional insight into variability in performance by stations). Results show that uncertainty is predominantly governed by model choice and station (location) effects, which together account for the majority of total variance, whereas the influence of NWP and BCM remains marginal. These findings highlight key vulnerabilities in operational inflow forecasting chains and provide actionable guidance for strengthening system resilience under real-time conditions.

KEYWORDS: uncertainty, forecast, machine learning, inflow, data-centric, hydrology

1 INTRODUCTION

The hydropower plants within the Drina and Lim River networks generate around 3295 GWh of electricity annually, contributing roughly 10% of Serbia's total electricity production (Web-1). These reservoirs play a crucial role in balancing the country's energy supply, especially during peak demand periods, as hydropower provides a flexible and renewable energy source. Efficient management of reservoir operations and water resources is essential for maintaining grid stability, optimizing generation schedules, and ensuring cost-effective electricity production. In addition, the Drina-Lim water system is vital for flood management and regional water security. High-flow events, which can be caused by extreme weather patterns, prolonged rainfall, or snowmelt, pose significant risks by potentially exceeding reservoir capacities and leading to flooding, endangering infrastructure, agricultural land, and local communities, resulting in substantial economic losses.

As climate change increases the frequency and severity of extreme hydrological events, the need for accurate and reliable streamflow forecasting has become more critical than ever. Advanced predictions of river flow and reservoir levels enable stakeholders to optimize hydropower production while simultaneously reducing flood risks. Machine learning (ML) models have shown significant capabilities for river flow prediction (Kratzert et al., 2018; Vinokić et al., 2025). Even though ML models can deliver highly accurate results efficiently, data inconsistencies can reduce their effectiveness (AL-Jarrah et al., 2015). Adopting a data-centric approach, we focus on the quality and reliability of model inputs, as well as the uncertainties they introduce into the system. By understanding these uncertainties and their cascade effects, hydropower operators gain deeper insights into informed decision-making. Furthermore, these insights serve as a blueprint for strengthening the system at its most vulnerable points, creating a more resilient forecasting system.

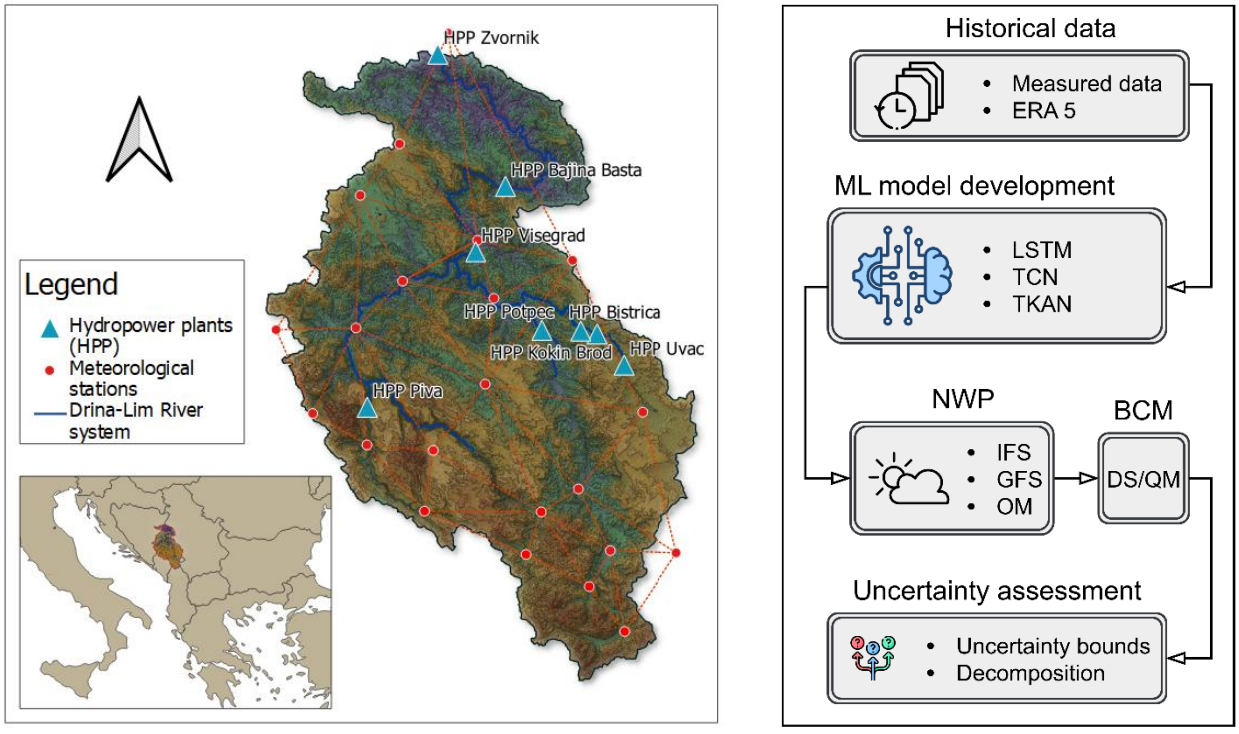
Our approach leverages global datasets, such as ECMWF ReAnalysis v5 (ERA5) (Web-2), as well as operational weather forecasts such as Integrated Forecasting System (IFS), Global Forecasting System (GFS), and Open Meteo (OM) weather forecast. Since these global numerical weather simulation models exhibit inherent biases at localized scales, two bias correction techniques are investigated: a pragmatic delta-scaling approach and empirical quantile mapping. To ensure robust predictions, three state-of-the-art machine-learning models—LSTM, TCN, and TKAN—are trained on pre-processed data, each offering unique perspectives in capturing hydrological patterns. A detailed uncertainty propagation analysis is conducted, dissecting each step to identify weak links and areas of high variability.

By systematically addressing key sources of uncertainty and integrating advanced AI-driven forecasting models, system operators and policymakers can make more informed decisions, enhance system resilience, and improve the long-term sustainability of energy and water resource management. This comprehensive approach highlights the critical role of uncertainty assessment in developing robust streamflow forecasting systems, ensuring efficient water resource utilization, infrastructure protection, and public safety amid escalating environmental challenges.

2 MATERIALS & METHODS

As mentioned in the previous section, the study area is the Drina-Lim river system (Figure 1a), located in Southeastern Europe, where the forecasting system is developed for total of 8 hydropower plants (HPP), where the unregulated reservoir inflow was estimated. This reservoir system is a multipurpose system, primarily used for hydropower generation, river regulation, and flood management. The catchment area of Drina river basin is characterized by predominantly mountain region, where rainfall-runoff process yields significant contributions to river flow dynamics and flood formation.

The general workflow of the study (Figure 1b) consists of five main steps: (1) preparing historical data for model training, (2) ML model development, (3) acquiring forcing data from different NWPs for the verification period, (4) implementing bias correction techniques, and (5) assessing uncertainty.



(a) (b)

Figure 1: (a) Case study and (b) study workflow

2.1 Data collection and processing

The dataset comprises daily time series data collected from 8 HPPs and 22 meteorological stations located in the Drina River basin, shown in more detail in Table 1. Data is structured in windowed features, consisting of several previous steps and forecast horizon. Lagged streamflow from previous 7 days is taken, along with precipitation and air temperature from 2 days before and 5 days ahead (referencing to forecast horizon). In training, this "future" data is taken from the existing observed dataset, while this data is taken from numerical weather predictions (NWPs) in inference.

Table 1: List of datasets used in the study

Type of data	Variable	Source
Hydrological data	Reservoir inflow data	EPS
Meteorological data	Precipitation sum	RHMSS (Web-3), ZHMS (Web-4)
	Air temperature	
Weather forecast data	Precipitation sum	IFS, GFS, UKMO
	Air temperature	IFS, GFS, UKMO

When ML models are operational and ready for real-time use, the weather forecast data must be transformed into daily time series data, and re-gridded onto the coordinates of meteorological measuring stations. This is achieved with spatial interpolation via inverse distance weighting (IDW) between a station location and the four nearest centres of the NWP grid. Equation (1) is used to determine the weights of neighboring center points, while Equation (2) calculates the interpolated value of a variable $\tilde{x}$ for a given x for 4 nearest neighbors:

$$w_j = \frac{{d_j}^{-\alpha}}{\sum_j {d_j}^{-\alpha}}; \quad (1)$$

$$\bar{x} = \sum_j w_j x_j; \quad (2)$$

where w_j represents weight of j-th center point, d_j its distance, and α denotes an order of an inverse distance (Atiah et al., 2023). For this study, inverse squared distance is used, which means that $\alpha = 2$.

Since the availability of data depends on location, models that are trained on specific HPPs utilize maximum available data (earliest time series begins in 1949 and latest in 1968, limited by dam construction / reservoir creation).

2.2 Bias correction methods (BCMs)

Two fundamentally different BCMs are used to estimate the impact of the use of bias correction in real-time operational conditions: delta scaling (DS), also known as delta change; and quantile mapping (QM). DS is a pragmatic method that utilizes, a principle of equating the mean monthly air temperature and precipitation sums. The strengths of this approach are its simplicity and cumulative volumetric consistency. However, its limiting factor is the assumption of linearity, effectively ignoring tails of distributions, which is of utmost importance for precipitation, as flattening the spike of heavy rainfall can produce underpredictions in river flow (Stojkovic & Simonovic, 2019). DS approach can be described as follows:

$$\Delta(\text{month}) = \begin{cases} \frac{\bar{x}_{obs}(\text{month})}{\bar{x}_{mod}(\text{month})}, & x = \text{precipitation} \\ \bar{x}_{obs}(\text{month}) - \bar{x}_{mod}(\text{month}), & x = \text{air temperature} \end{cases} \quad (3)$$

$$\hat{x}_{mod} = \begin{cases} \Delta * x, & x = \text{precipitation} \\ x + \Delta, & x = \text{air temperature} \end{cases} \quad (4)$$

where *obs* subscript represent observed, and *mod* modelled values, $\bar{x}$ denotes mean value of a variable, and $\hat{x}$ a bias corrected value. Quantile mapping (QM), on the other hand, suits precipitation data better, where cumulative distribution function (CDF) is mapped onto the reference CDF, correcting the entire probability distribution (Stojkovic & Simonovic, 2019). This is especially useful in situations with extreme events, such as extreme precipitation because of the high sensitivity to such occurrences (Vinokić et al., 2023). QM is generally expressed through following function:

$$\hat{x}_{mod}(q) = F_{obs}^{-1}(F_{mod}[x_{mod}(q)]). \quad (5)$$

Transfer functions are calculated from the only available period where NWP and observed values overlap, which is from 2022, limiting the period from which a transfer function can be extracted.

2.3 Machine Learning models

Three state-of-the-art ML model architectures are used, including Long Short-Term Memory (LSTM), Temporal Convolutional Network (TCN), and Temporal Kolmogorov Arnold Network (TKAN).

LSTMs (Hochreiter & Schmidhuber, 1997) augment traditional recurrent neural networks by incorporating a memory cell regulated by input, forget, and output gates. These components enable effective preservation and updating of long-term temporal information, making LSTMs particularly suitable for modelling hydrological phenomena with delayed responses, such as runoff generation following precipitation (Kratzert et al., 2018). TCNs (Lea et al., 2016) model sequences using 1D dilated causal convolutions, allowing them to capture long-range dependencies while maintaining strict temporal causality. Residual connections facilitate the training of deep models, mitigating vanishing gradients. TCNs have shown strong performance in time-series forecasting due to their ability to model both short-term and long-term temporal patterns (Bai et al., 2017). TKANs extend Kolmogorov–Arnold Networks (KANs), which replace fixed activation functions with learnable univariate functions (often splines) applied on edges rather than nodes (Liu et al., 2024). By modifying KANs to incorporate temporal dependencies, TKANs adapt this flexible functional representation to sequence modelling. Their ability to learn expressive nonlinear transformations with fewer parameters makes them a promising emerging architecture for hydrological time-series forecasting (Vinokić et al, 2025).

ML models are trained on data from different periods, from the earliest 1949, to the latest 1968, to the year 2014, while the following two years used for validation, and 2017 and 2018 for testing. However, for the testing of operational conditions an estimation of uncertainty in real-time forecasting scenarios, a verification set is selected to be year 2023, as the only remaining year with available data, after using pre previous one for extracting the transfer functions.

2.4 Uncertainty decomposition

The uncertainty bounds are derived from an ensemble of 27 model predictions. Each prediction differs from the others in terms of the numerical weather prediction (NWP) forcing, bias-correction method (BCM), or machine-learning (ML) model employed. This structured ensemble enables the decomposition of predictive uncertainty across these three sources by treating them as independent factors. The uncertainty decomposition is performed using factorial ANOVA, as illustrated in Figure 2.

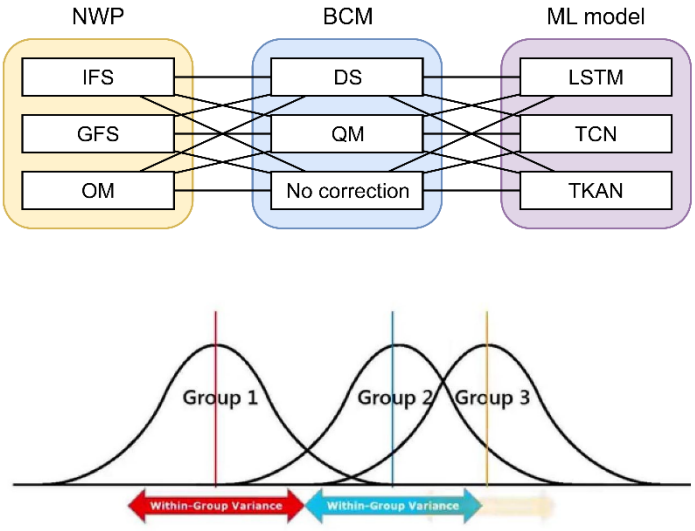


Figure 2: Uncertainty decomposition scheme

The ANOVA-derived variance corresponds to interpretable uncertainty contributions for each of the three factors, where the proportion of total explained variance by a given factor R_i is computed as follows:

$$R_i = \frac{v_i}{v_{total}} \times 100, \quad (6)$$

ANOVA-based uncertainty decomposition is run on multiple scalar metrics to capture different dimensions of model behaviour, together providing a comprehensive decomposition across factors, by averaging their respective results.

2.5 Performance metrics

ANOVA is applied to scalar performance metric or a scalar characteristic, computed for every factor combination, indicating model performance. Since the analysis is performed on multiple different HPPs, the relative or normalized metrics are applied for comparability across stations, as shown in Table 2.

Table 2: Overview of used performance metrics

Performance metric	Equation	Use
Normalized Root Mean Squared Error (NRMSE)	$NRMSE = \frac{1}{n}\sqrt{\frac{\sum_i^n \left(Q_i^{mod} - Q_i^{obs}\right)^2}{\overline{Q}^{obs}}}$	Skill metrics for ANOVA
Nash-Sutcliffe Efficiency (NSE)	$NSE = \frac{\sum_i^n \left(Q_i^{obs} - Q_i^{mod}\right)^2}{\sum_j^n \left(Q_i^{obs} - \overline{Q}^{obs}\right)^2}$	
Prediction Interval Coverage Percentage (PICP)	$PICP = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}(y_i \in [\hat{Q}_i^{\,L}, \hat{Q}_i^{\,U}])$	Uncertainty bounds

* Q_i^{mod} and Q_i^{obs} represent the predicted and measured discharge for i-th time-step respectfully, $\overline{Q}^{obs}$ denotes the mean measured river flow for the period, while n denotes the number of time-steps. $\hat{Q}_i^{\,L}$ and $\hat{Q}_i^{\,U}$ represent the lower and upper value of uncertainty bounds for a single time-step.

3 RESULTS & DISCUSSION

Ensemble forecasting provides a significant insight into uncertainty, while also providing reliable forecasting bounds. Depending on the station, the prediction envelope covers from 65% to 93% observed values within the 10% margin of error, measured with PICP. Note that reservoir inflow estimates can be inaccurate, impacted with the reservoir system operation. Therefore, heavily regulated inflows, such as on HPP Zvornik or HPP Bajina Bašta, where PICP was the lowest (65% and 70%), can be difficult to estimate natural flow, resulting in somewhat noisy data, especially in low flow periods. Ensemble forecast in verification (operational-like) setting are shown in Figure 3. The distribution of uncertainty is quite different for each of the stations, indicating a high level of uncertainty is connected to catchment-specifics and river inflows that are not considered as a factor, since there is only one source of inflow data. Upon analyzing the results that include HPPs (location) as a factor, the variability attributed to station factor is 31% to 65% depending on chosen metric, confirming that the large portion of the uncertainty spread is location-based.

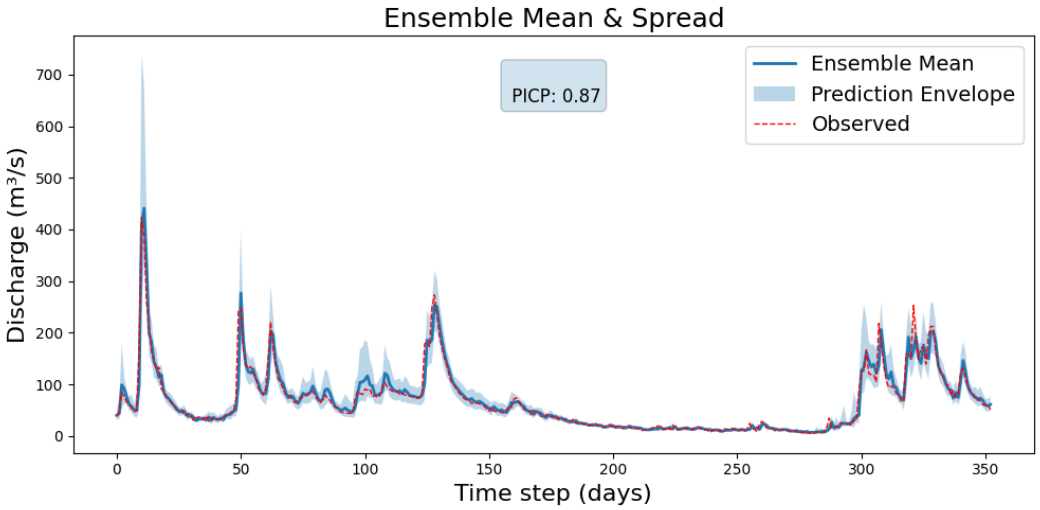


Figure 3: Example of ensemble forecasting with uncertainty bounds on the verification set.

As shown in Table 3, model choice emerges as the dominant contributor to total uncertainty, explaining between 57% and 96% of the overall variance. By comparison, NWP and BCM selections exhibit mostly marginal influence, under 10% combined for the majority of HPPs, while the unexplained variance spans from 3% to 31%.

Table 3: Uncertainty contributions for each reservoir inflow forecast averaged for three skill scores

Factor	Uvac	Kokin Brod	Bistrica	Piva	Potpeć	Višegrad	Bajina Bašta	Zvornik
NWP	4.2%	3.8%	11.6%	3.1%	0.7%	0.3%	1.1%	0.5%
BCM	1.1%	3.6%	2.6%	14.5%	13.1%	1.0%	2.0%	1.5%
ML	81.4%	61.8%	64.5%	57.4%	66.5%	95.5%	91.7%	94.8%
Residual	13.3%	30.8%	21.3%	25.0%	19.7%	3.1%	5.2%	3.2%

4 CONCLUSION

This study confirms that uncertainty in ML-based daily reservoir inflow forecasting on the Drina-Lim system is dominated by model structure and location-specific characteristics, particularly under strong hydraulic regulation. While ensemble forecasts provide reliable predictive bounds, a substantial share of uncertainty remains inherently station-dependent, highlighting the critical role of local hydrological and operational influences.

When considering station-specific models, the largest share of the explained variance is attributed to the ML model architecture. This is expected given the design of this study, where variability is introduced primarily through the operational setting rather than the training process. The absence of variability in training data and optimization constrains the contribution of other uncertainty sources, thereby emphasizing structural differences between model types. These findings highlight a clear direction for future research, where variability should be systematically introduced at the training stage by employing substantially larger ensembles of models trained on different data realizations, architectures, and hyperparameter configurations.

5 ACKNOWLEDGEMENTS

The research presented herein received financial support from the European Union's Horizon Europe project ARTIFACT, under Grant Agreement 101159480. We also extend our gratitude to Joint Stock Company Elektroprivreda Srbije for their trust and support in the project "Prediction of Daily Reservoir Inflows at the Drina–Lim Hydropower Plants Using Artificial Intelligence Techniques." Their collaboration and funding were instrumental in enabling the research activities and achieved outcomes.

REFERENCES

Al-Jarrah, O. Y., Yoo, P. D., Muhaidat, S., Karagiannidis, G. K., & Taha, K. (2015). Efficient machine learning for big data: A review. Big Data Research, 2(3), 87-93.

C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal Convolutional Networks for Action Segmentation and Detection," Nov. 16, 2016, arXiv: arXiv:1611.05267. Accessed: Jul. 14, 2024. [Online]. Available: http://arxiv.org/abs/1611.05267.

F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, "Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks," Hydrol. Earth Syst. Sci., vol. 22, no. 11, pp. 6005–6022, Nov. 2018, doi: 10.5194/hess-22-6005-2018.

S. Ahmad and S. P. Simonovic, "System Dynamics Modeling of Reservoir Operations for Flood Management," J. Comput. Civ. Eng., vol. 14, no. 3, pp. 190–198, Jul. 2000, doi: 10.1061/(ASCE)0887-3801(2000)14:3(190).

S. Bai, J. Z. Kolter, and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," Apr. 19, 2018, arXiv: arXiv:1803.01271. Accessed: Jul. 14, 2024. [Online]. Available: http://arxiv.org/abs/1803.01271

S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

Stojkovic, M., & Simonovic, S. P. (2019). Mixed General Extreme Value Distribution for Estimation of Future Precipitation Quantiles Using a Weighted Ensemble - Case Study of the Lim River Basin (Serbia). Water Resources Management.

Vinokić, L., Dotlić, M., Prodanović, V., Kolaković, S., Simonovic, S. P., & Stojković, M. (2025). Effectiveness of three machine learning models for prediction of daily streamflow and uncertainty assessment. *Water Research X*, *27*, 100297.

Vinokić, L., Stojković, M., Kolaković, S. (2023). Bias correction and weighting methods to shape precipitation under the climate change options. Proceedings of the International Conference Synergy of Architecture & Civil Engineering. SINARG 2023. Niš, Serbia.

W. A. Atiah et al., "Bias correction and spatial disaggregation of satellite-based data for the detection of rainfall seasonality indices," Heliyon, vol. 9, no. 7, p. e17604, Jul. 2023, doi: 10.1016/j.heliyon.2023.e17604.

Z. Liu et al., "KAN: Kolmogorov-Arnold Networks," Jun. 16, 2024, arXiv: arXiv:2404.19756. Accessed: Jul. 14, 2024. [Online]. Available: http://arxiv.org/abs/2404.19756

Web sites:

Web-1: https://www.eps.rs/cir/dlhe/Pages/default.aspx, consulted 10 December, 2025

Web-2: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-complete, consulted 15 July, 2024.

Web-3: https://www.hidmet.gov.rs/index_eng.php, consulted 11 September 2025.

Web-4: https://www.meteo.co.me/, consulted 15 July, 2024.