

Application of Machine Learning for Runoff Prediction

Subjects: Water Resources

Contributor: Ze Liu, Jingzhao Zhou, Xiaoyang Yang, Zechuan Zhao, Yang Lv

Water resource modeling is an important means of studying the distribution, change, utilization, and management of water resources. By establishing various models, water resources can be quantitatively described and predicted, providing a scientific basis for water resource management, protection, and planning. Traditional hydrological observation methods, often reliant on experience and statistical methods, are time-consuming and labor-intensive, frequently resulting in predictions of limited accuracy. However, machine learning technologies enhance the efficiency and sustainability of water resource modeling by analyzing extensive hydrogeological data, thereby improving predictions and optimizing water resource utilization and allocation.

Keywords: water resource ; machine learning ; runoff prediction ; medium- and long-term runoff ; short-term runoff

1. Introduction

With the escalating severity of global water shortage problems, the management and efficient utilization of water resources have become a critical research focus ^{[1][2]}. In China, the total volume of freshwater resources amounts to 2800 billion cubic meters, representing 6% of global water resources and ranking fourth globally, following Brazil, Russia, and Canada ^[3]. However, China's per capita water resources stand at only 2300 cubic meters, equivalent to just a quarter of the global average, categorizing it among the countries with the scarcest per capita water resources globally ^[4]. In China, precipitation diminishes from the southeast coast to the northwest inland, categorized into five zones: rainy, humid, semi-humid, semi-arid, and arid ^[5]. Owing to the heterogeneous distribution of precipitation across regions, a pronounced supply–demand imbalance of water resources exists in China, particularly in the northern regions ^[6]. With population growth and economic advancement, the demand for water resources steadily increases, while the supply remains limited, leading to water shortages in some areas ^[7]. In some regions, water resource utilization efficiency in China is noticeably low, primarily due to inadequate focus on water resources, insufficient effective water-saving measures, and a lack of institutional guarantees ^[8]. Furthermore, with the accelerated pace of industrialization and urbanization, substantial quantities of industrial and agricultural wastewater, domestic sewage, and solid waste are persistently discharged, leading to severe pollution of rivers, lakes, and groundwater ^[9]. Consequently, there is a critical need to enhance the protection and utilization of water resources and to implement effective measures aimed at improving water resource utilization efficiency ^[10].

Water resources constitute renewable resources and demonstrate variability in annual and interannual quantities, possessing a distinct cycle and pattern (**Figure 1**). Influenced by solar radiation, Earth's gravity, and other physical processes, including evapotranspiration, precipitation, soil infiltration, surface runoff, and underground flow, water translocates from one location to another ^[11]. Water resources form a critical foundation for the survival and development of human society ^[12]. To enhance the management and utilization of water resources, various computational models have been employed in water resource management ^[13]. The traditional hydrological model operates on the principle of the hydrological cycle, articulating the components of the hydrological cycle system ^[14]. These models hold significant physical importance and are straightforward to elucidate; however, their development process is complex and requires extensive expertise from developers ^[15]. Machine learning can process massive amounts of data, extract valuable information, and automatically construct models to predict future trends ^[16]. In water resource management, such data enable a more comprehensive consideration of problems and the establishment of improved decision-making models ^[17]. Consequently, machine learning, owing to its robust data mining capabilities, is widely applied in areas like water resource supply and demand prediction, flood risk management, water quality monitoring, and forecasting scenarios ^{[18][19][20][21][22]}.

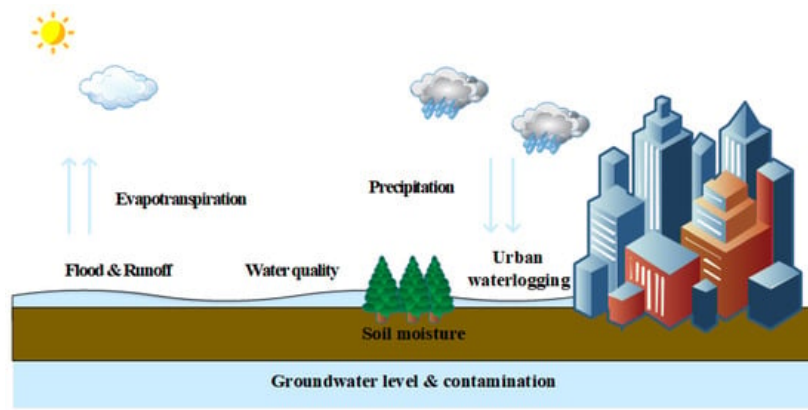


Figure 1. Schematic diagram of water resource cycle.

Forecasting runoff plays a crucial role in optimizing water resource systems and mitigating the impacts of destructive natural disasters, such as floods and droughts, through both long-term planning and short-term emergency warnings [23][24][25]. Due to the complexity of the causes of runoff and the challenging nature of understanding their mechanisms, constructing machine learning-based models for runoff prediction emerges as an effective solution [26]. When classified by time scale, runoff prediction can be segmented into medium- and long-term predictions, as well as short-term forecasting [27].

2. Medium- and Long-Term Runoff Prediction

Prediction of medium- and long-term runoff is vital for developing water resource scheduling plans that span extended durations, thereby significantly impacting water resource management [28]. Ghumman et al. developed an ANN-based runoff model using continuously measured monthly rainfall and runoff data, applying it to predict monthly runoff in the Hoab River Basin of Pakistan [29]. ANNs do not require a detailed investigation of hydrological and geological parameters of the catchment to perform similarly to traditional conceptual models. Nevertheless, the dataset frequently contains a high proportion of noise and errors, which can impede the ability of ANN-based models to make efficient and accurate predictions. Owing to climate change and human activities, natural runoff often contains multiple frequency components, posing a challenge for traditional ANN-based models in efficiently capturing the underlying change processes.

Tan et al. employed multi-year runoff data and introduced an ANN-based runoff model combined with ensemble empirical modal decomposition (EEMD). The results demonstrate that the proposed EEMD-ANN model significantly enhances the accuracy of the ANN-based method in predicting the medium- and long-term runoff time series [30]. Liao proposed a hybrid framework for long-term runoff prediction, incorporating pre-inflow and specific meteorological factors, such as precipitation, evapotranspiration, solar radiation, soil temperature, etc., as input features. The results showed that the accuracy is improved by combining EEMD and ANN for modeling [31]. The challenge in medium- and long-term runoff prediction lies in its low accuracy, arising from the extended prediction period and the complexity of the runoff genesis mechanism. Consequently, identifying key factors is of critical importance. Han et al. introduced a LSTM-based model, AT-LSTM, combining double attention mechanisms in the input and hidden layers. The model uses rainfall, potential evapotranspiration, and monthly climate phenomenon index data (including 88 atmospheric circulation indicators, 26 sea surface temperature indicators, and 16 other indicators) as inputs for long-term runoff prediction. The results demonstrate that the AT-LSTM model effectively enhances the accuracy of long-term prediction and identifies the dynamic effects of the input factors [32].

3. Short Term Runoff Prediction

Zealand et al. examined the efficacy of ANNs for short-term runoff prediction, and the results demonstrating that the ANN-based models consistently outperform traditional models [33]. Kratzert utilized meteorological forcing data (maximum air temperature, precipitation, radiation, vapor pressure, etc.) and static catchment characteristics (drought index, PET mean, max water content, geological permeability, forest fraction, etc.) as features to build an LSTM-based model for daily runoff prediction. The obtained results surpassed those of the well-established physical model [23]. Based on data from 98 rainfall runoff events, Hu et al. compared the performance of the ANN-based and LSTM-based models for simulating runoff processes during runoff events over a delivery period of 1 to 6 h. The results showed that the LSTM-based model outperformed the ANN-based models [34]. Gao et al. developed a short-term runoff prediction model using LSTM and gated recurrent unit (GRU) networks, based on hourly flow measurements from one runoff station and hourly rainfall data

from four rainfall stations [35]. The experiments demonstrated that the GRU model requires the least training time and has a simpler structure, making it the preferred method for short-term runoff prediction.

Despite achieving high accuracy in runoff prediction, no model can guarantee maximum certainty in predictions due to issues like noisy or incomplete data. To improve prediction accuracy and reduce data dependency, Naganna et al. utilized the runoff time series of the Gauvery River in India. They applied deep learning techniques (CNN, RF, and Gradient Tree Boosting (GTB)), incorporating the Information Criterion (AIC) and Bayesian Information Criterion (BIC) to select the ideal input parameters for predicting the daily scale of multiple basins in the Gauvery River in India. The results indicated that deep learning (CNN), combined with AIC and BIC to select the ideal inputs to the model, achieves excellent prediction accuracy [36].

In summary, several features influence runoff, with rainfall, evapotranspiration, temperature, and radiation being the most crucial factors. For medium- and long-term runoff prediction, the best model is the AT-LSTM model combining double attention mechanisms, while for short-term runoff prediction, the CNN-based model combining AIC and BIC is the most effective. Presently, global climate change has a series of impacts on the hydrological cycle, subsequently influencing the hydrological processes in the watershed. The non-stationary hydrological sequences, influenced by factors like climatic and meteorological conditions, subsurface conditions, and human activities, introduce new challenges to hydrological forecasting and other research work [37]. Under non-stationary conditions, traditional runoff forecasting methods cannot be directly applied. In the future, it is necessary to explore the physical mechanism, integrate machine learning methods, and incorporate hydrological–meteorological information to establish a runoff prediction model suitable for non-stationary conditions.

References

1. Ren, J.; Zhao, D. Recent Advances in Reticular Chemistry for Clean Energy, Global Warming, and Water Shortage Solutions. *Adv. Funct. Mater.* 2023, 5, 2307778.
2. Gharib, A.A.; Blumberg, J.; Manning, D.T.; Goemans, C.; Arabi, M. Assessment of vulnerability to water shortage in semi-arid river basins: The value of demand reduction and storage capacity. *Sci. Total. Environ.* 2023, 871, 161964.
3. Yang, J.; Yang, P.; Zhang, S.; Wang, W.Y.; Cai, Y.; Hu, S. Evaluation of water resource carrying capacity in the middle reaches of the Yangtze River Basin using the variable fuzzy-based method. *Environ. Sci. Pollut. Res.* 2022, 30, 30572–30587.
4. Li, Y.; Han, Y.; Liu, B.; Li, H.; Du, X.; Wang, Q.; Wang, X.; Zhu, X. Construction and application of a refined model for the optimal allocation of water resources—Taking Guantao County, China as an example. *Ecol. Indic.* 2023, 146, 109929.
5. Li, Z.; Liu, H. Temporal and spatial variations of precipitation change from Southeast to Northwest China during the period 1961–2017. *Water* 2020, 12, 2622.
6. Guo, Y.; Shen, Y. Agricultural water supply/demand changes under projected future climate change in the arid region of northwestern China. *J. Hydrol.* 2016, 540, 257–273.
7. Zhang, Y.; Khan, S.U.; Swallow, B.; Liu, W.; Zhao, M. Coupling coordination analysis of China's water resources utilization efficiency and economic development level. *J. Clean. Prod.* 2022, 373, 133874.
8. Zhang, H.; Jin, G.; Yu, Y. Review of river basin water resource management in China. *Water* 2018, 10, 425.
9. Lin, L.; Yang, H.; Xu, X. Effects of water pollution on human health and disease heterogeneity: A review. *Front. Environ. Sci.* 2022, 10, 880246.
10. Makanda, K.; Nzama, S.; Kanyerere, T. Assessing the role of water resources protection practice for sustainable water resources management: A review. *Water* 2022, 14, 3153.
11. Loucks, D.P. Sustainable water resources management. *Water Int.* 2000, 25, 3–10.
12. Vorosmarty, C.J.; Green, P.; Salisbury, J.; Lammers, R.B. Global water resources: Vulnerability from climate change and population growth. *Science* 2000, 289, 284–288.
13. Wu, B.; Zheng, Y.; Wu, X.; Tian, Y.; Han, F.; Liu, J.; Zheng, C.B. Optimizing water resources management in large river basins with integrated surface water-groundwater modeling: A surrogate-based approach. *Water Resour. Res.* 2015, 51, 2153–2173.
14. Wang, J.; Shi, P.; Jiang, P.; Hu, J.; Qu, S.; Chen, X.; Chen, Y.; Dai, Y.; Xiao, Z. Application of BP neural network algorithm in traditional hydrological model for flood forecasting. *Water* 2017, 9, 48.

15. Shen, H.; Tolson, B.A.; Mai, J. Time to update the split-sample approach in hydrological model calibration. *Water Resour. Res.* 2022, 58, e2021WR031523.
16. Rani, K.S.; Kumari, M.; Singh, V.B.; Sharma, M. Deep learning with big data: An emerging trend. In *Proceedings of the 2019 19th International Conference on Computational Science and Its Applications (ICCSA)*, Saint Petersburg, Russia, July 1–4 2019; pp. 93–101.
17. Anjum, R.; Parvin, F.; Ali, S.A. Machine Learning Applications in Sustainable Water Resource Management: A Systematic Review. In *Emerging Technologies for Water Supply, Conservation and Management* Springer Water; Springer: Cham, Switzerland, 2023; pp. 29–47.
18. Mekanik, F.; Imteaz, M.A.; Gato-Trinidad, S.; Elmahdi, A. Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes. *J. Hydrol.* 2013, 503, 11–21.
19. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep learning and process understanding for data-driven Earth system science. *Nature* 2019, 566, 195–204.
20. Uc-Castillo, J.L.; Marín-Celestino, A.E.; Martínez-Cruz, D.A.; Tuxpan-Vargas, J.; Ramos-Leal, J.A. A systematic review and meta-analysis of groundwater level forecasting with machine learning techniques: Current status and future directions. *Environ. Model. Softw.* 2023, 168, 5788.
21. Arrighi, C.; Castelli, F. Prediction of ecological status of surface water bodies with supervised machine learning classifiers. *Sci. Total Environ.* 2023, 857, 159655.
22. Ghobadi, F.; Kang, D. Application of Machine Learning in Water Resources Management: A Systematic Literature Review. *Water* 2023, 15, 620.
23. Kratzert, F.; Klotz, D.; Herrnegger, M.; Sampson, A.K.; Hochreiter, S.; Nearing, G.S. Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resour. Res.* 2019, 55, 11344–11354.
24. Fang, J.J.; Yang, L.; Wen, X.H.; Li, W.; Yu, H.; Zhou, T. A deep learning-based hybrid approach for multi-time-ahead streamflow prediction in an arid region of Northwest China. *Hydrol. Res.* 2024, nh2024124.
25. Liu, J.; Ren, K.; Ming, K.; Qu, J.; Guo, W.; Li, H. Investigating the effects of local weather, streamflow lag, and global climate information on 1-month-ahead streamflow forecasting by using XGBoost and SHAP: Two case studies involving the contiguous USA. *Acta Geophys.* 2023, 71, 905–925.
26. Yao, Z.; Wang, Z.; Wang, D.; Wu, J.; Chen, L. An ensemble CNN-LSTM and GRU adaptive weighting model based improved sparrow search algorithm for predicting runoff using historical meteorological and runoff data as input. *J. Hydrol.* 2023, 625, 129977.
27. Liu, Y.; Wang, L.; Yang, L.; Liu, X.; Wang, L. Runoff Prediction and Analysis Based on Improved CEEMDAN-OS-QR-ELM. *IEEE Access* 2021, 9, 57311–57324.
28. Yang, M.X.; Wang, H.; Jiang, Y.; Lu, X.; Xu, Z.; Sun, G. GECA Proposed Ensemble-KNN Method for Improved Monthly Runoff Forecasting. *Water Resour. Manag.* 2020, 34, 849–863.
29. Ghumman, A.R.; Ghazaw, Y.M.; Sohail, A.R.; Watanabe, K. Runoff forecasting by artificial neural network and conventional model. *Alex. Eng. J.* 2011, 50, 345–350.
30. Tan, Q.F.; Lei, X.H.; Wang, X.; Wang, H.; Wen, X.; Ji, Y.; Kang, A.Q. An adaptive middle and long-term runoff forecast model using EEMD-ANN hybrid approach. *J. Hydrol.* 2018, 567, 767–780.
31. Liao, S.L.; Wang, H.; Liu, B.X.; Ma, X.; Zhou, B.; Su, H. Runoff Forecast Model Based on an EEMD-ANN and Meteorological Factors Using a Multicore Parallel Algorithm. *Water Resour. Manag.* 2023, 37, 1539–1555.
32. Han, D.Y.; Liu, P.; Xie, K.; Li, H.; Xia, Q.; Cheng, Q.; Xia, J. An attention-based LSTM model for long-term runoff forecasting and factor recognition. *Environ. Res. Lett.* 2023, 18, 024004.
33. Zealand, C.M.; Burn, D.H.; Simonovic, S.P. Short term streamflow forecasting using artificial neural networks. *J. Hydrol.* 1999, 214, 32–48.
34. Hu, C.H.; Wu, Q.; Li, H.; Jian, S.; Li, N.; Lou, Z. Deep Learning with a Long Short-Term Memory Networks Approach for Rainfall-Runoff Simulation. *Water* 2018, 10, 1543.
35. Gao, S.; Huang, Y.; Zhang, S.; Han, J.; Wang, G.; Zhang, M.; Lin, Q. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. *J. Hydrol.* 2020, 589, 125188.
36. Naganna, S.R.; Sreedhara, B.; Muttana, S.; Marulasiddappa, S.B.; Balreddy, M.S.; Yaseen, Z.M. Daily scale streamflow forecasting in multiple stream orders of Cauvery River, India: Application of advanced ensemble and deep learning models. *J. Hydrol.* 2023, 626, 130320.
37. Zhang, Y.; Zheng, H.; Zhang, X.; Leung, L.R.; Liu, C.; Zheng, C.; Blöschl, G. Future global streamflow declines are probably more severe than previously estimated. *Nat. Water* 2023, 1, 261–271.

