Impact of Climate Variables on Reservoir Outflow: An AI Approach; A Case Study of Khanpur Reservoir, Pakistan

Afzal Ahmed[1]*, Muhammad Talha Maqsood[1], Ghufran Ahmed Pasha[1], Maaz Athar[1], Shahryar Javed[1], Waleed Ahmed[1]

[1] Civil Engineering Department, University of Engineering & Technology Taxila 47050, Rawalpindi, Pakistan; afzal.ahmed@uettaxila.edu.pk, 21-ce-78@students.uettaxila.edu.pk , ghufran.ahmed@uettaxila.edu.pk , 21-ce-46@students.uettaxila.edu.pk , 21-ce-70@students.uettaxila.edu.pk , 21-ce-166@students.uettaxila.edu.pk

***** Correspondence: afzal.ahmed@uettaxila.edu.pk

Abstract

This research examines the effect of climate factors on the outflow of the Khanpur Reservoir in Pakistan by employing ANN for modeling. To forecast the reservoir outflow, the hydrological and climate variables such as seepage, maximum temperature, rainfall, live storage, evaporation, humidity and inflow are utilized in the analysis. ANN models with 5x5 (5 neurons in each two layers), 10x10, 15x15, 20x20, 25x25 neurons combination were created and evaluated, and it was found that the model with 20 neurons in each hidden layer had the best performance having R^2 value of the training and validation set was 0.95 and 0.85 respectively. The sensitivity analysis through interaction profiles revealed that live storage is the most significant predictor of reservoir outflow due to its high main effect (84410) and comparatively low interaction with other variables (total effect: 9790). This underscores the importance of accurately measuring and incorporating live storage in reservoir management strategies. Moreover, inflow had moderate direct impact (0.097) and the highest overall impact (0.423) revealing it as another most significant and sensitive predictor of outflow. Rainfall exhibited minimal direct influence with main effect: 0.029) but significant interaction effects (total effect: 0.411) showing its moderate effects on outflow. Total seepage and maximum temperature had small main effects: 0.021 and 0.038, respectively, but their moderate total effects were 0.367 and 0.341, respectively, pointed to their indirect significance. In contrast, evaporation and relative humidity exhibited very small individual and combined impacts on outflow (main effect: 0 and 0.014; total effect: 29.03 and 0.129, respectively). These results reveal the strong impact of climate variables on reservoir outflow prediction and capability of ANN in modelling these aspects. This research adds to the existing literature on employing AI and machine learning for optimizing reservoir management and advancing water management techniques under the new climate reality.

Keywords: Reservoir; Artificial neural network; Outflow; Khanpur Dam; Water management

1. Introduction

Reservoirs are basic facilities in water resource system, and they play important roles including flood management, water supply, irrigation, and generation of hydroelectric power. Climatic and hydrological factors play a very central role in the operation of a reservoir where the inflow into the reservoir is determined by rainfall. Forecasting of the inflow is crucial in determining the availability of water in reservoirs and in planning how best to utilize water resources in a given society, particularly in arid zones where water is a very scarce resource [1][2]. These forecasts are, however, contrary to the complexities of the interactions between climatic variables such as precipitation, temperature, and evaporation, and the hydrologic processes that control reservoir inflows [3]. In addition, agricultural consumption, industrial consumption, and hydropower production make reservoir operation more challenging [4].

Semi-distributed and physically based models have been used in the past to predict inflows and outflows to manage reservoirs. These models such as the Hydrologic Engineering Centre's Hydrologic Modelling System (HEC-HMS) have however been used because they are able to capture certain physical processes in the catchment area [5][6]. But they encounter challenges of being data-intensive, computationally expensive and requiring precise calibration most of the time. Besides, the forecast accuracy of these models can be weakened depending on the volume of data for calibration and nonlinear characteristics of the catchment area triggered by climate change and changes in land use [7][8].

To overcome these challenges, data based models especially Artificial Neural Network (ANN) models are increasingly used in hydrological forecasting because of their capability of modelling non-linear relationships with less data demands [9][10]. ANN models have been used in several studies to estimate inflow and streamflow, in addition to using them to control reservoirs under different climates [11][12]. However, the performance of the ANN models can be improved by carefully selecting the inputs and combining the hybrid modelling approach that incorporates physical and data-driven models [13][14].

Based on these developments, the current research examines the application of ANN models for predicting reservoir outflow at Khanpur Reservoir, Pakistan, in different climatic conditions. This research utilizes daily data of five years to assess the efficiency of various ANN architectures, including models with two hidden layers and different combinations of neurons in each layer, for estimating reservoir outflow and input variables, seepage, maximum temperature, rainfall, live storage, evaporation, humidity and inflow. According to the results of this study, the ANN models can forecast the highest level of accuracy on the training set and validation set with the R^2 of 0.95 and 0.85, respectively when the architecture of model is 20x20 (20 neurons in both layers). The results also show that ANNs can be used as a viable tool in reservoir management, especially in data scarce areas. This research falls within the line of research that incorporate AI and hydrological modelling, and provides a review on the application of ANN based approaches for real time outflow forecasting in semi arid regions[15][16][17].

2. Study Area and Data

Khanpur Dam is constructed on the Khanpur channel and is located near the town of Khanpur in the Rawalpindi district of Punjab, Pakistan at a geographical coordinate of 33°48′25″ N latitude and 72°56′10″ E longitude as depicted in Figure 1. The reservoir has its catchment area at an altitude of 457 metres above the sea level and the gross storage capacity of the reservoir is 1.58 MAF. The dam is important in providing water for water for agriculture, domestic use and for control of floods in the neighboring areas.

Figure 1. Location map of Khanpur dam

3. Methodology

3.1. Data Collection

The data used in this study was collected from the Khanpur Reservoir in Pakistan which consist of daily data for 5 years. The dataset has seepage, maximum temperature, rainfall, live storage, inflow as inputs and outflow as output for the climatic and hydrological model. To ensure that the data was clean the outliers were removed from the data set. Some observations that deviate from the rest were removed by observing the graphs of the output variable (outflow) against the input variables.

Before feeding it to ANN, the independent variables were normalised in a way that all the inputs are in the similar range as it is beneficial for the model. Any gaps in the data were filled through interpolation methods to ensure the quality of the dataset is preserved.

3.2. ANN Model in JMP

To design the artificial neural network model JMP was used to design and set up the neural network. The structure of the ANN model and the number of neurons in the model, input, hidden and output layer is depicted in the Figure 2. In this study, the ANN architecture consisted of two hidden layers, with neurons varying across five different configurations: 5×5, 10×10, 15×15, 20×20 and 25×25. The activation function used for both the above layers was TanH that can easily address non linear relationships. The penalty method used was squared which reduces large weight values and hence guard against overfitting.

The performance of the training and validation of the model was assessed using R^2, RMSE and SSE. In order to cheque generalization abilities of the model, the holdback value was set to 0.33, which means that 33% of the data was used for the validation and 67% for the model training. This approach can help assess the model on other data, thus preventing overfitting and improving its accuracy.

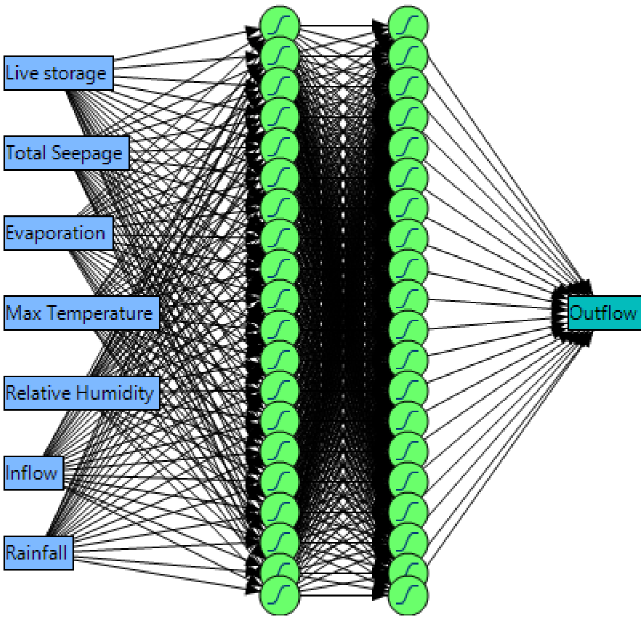


Figure 2. ANN Diagram Showing Inputs, Outputs, and Neurons

3.3. Model Evaluation and Performance Metrics

After that the model was evaluated on the test data set which has not been used during the model training procedure. In order to evaluate the model predictiveness of the outflow, the R^2 value for both training and validation groups was measured. The training R^2 was 0.95 and validation R^2 was 0.85 which depicts that the current model is more appropriate to the training data set and reasonable on the validation data set.

4. Results and Discussion

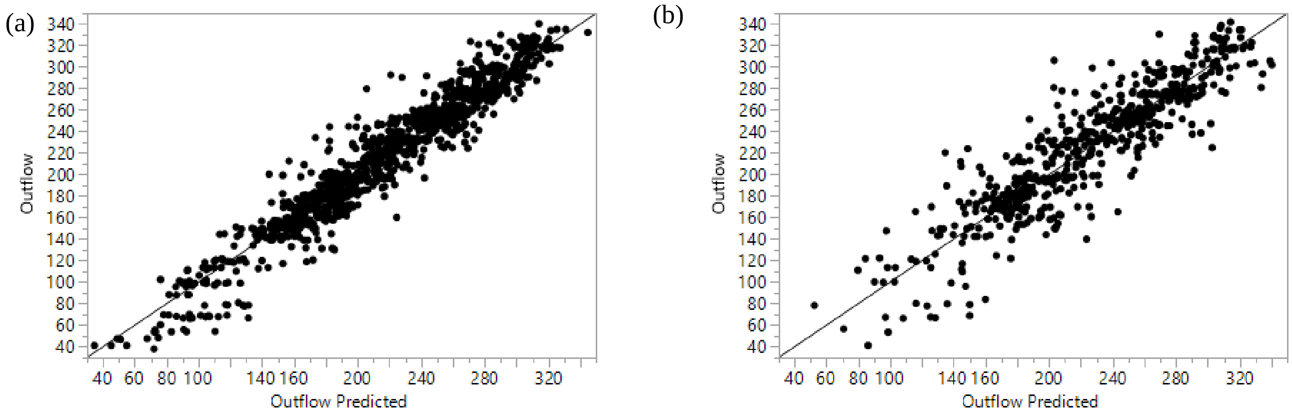
4.1. Model Performance

Several measures of performance were used to test the developed ANN model in JMP. It was 0.95 for the training data set and 0.85 for the validation set, which testifies high predictive power of the model. The findings of these analyses imply that the model fits well on the training data and also has promising capability to perform well on other data sets.

The RMSE and SSE for the training and validation data were also analysed to provide support for the model. The low values of RMSE and SSE suggest that the proposed model accurately predicts outflow values which will be useful in reservoir management.

4.2. Prediction Performance

In order to compare the ANN model's accuracy, graphs of predicted vs observed outflow were plotted. These graphs (Figure 3 (a), (b)) show how well the model fits the outflow data across the whole data range. The actual and predicted outflow values also had a near perfect positive correlation in that the points were clustered closely to the 45-degree line proving that the model had high prediction precision.

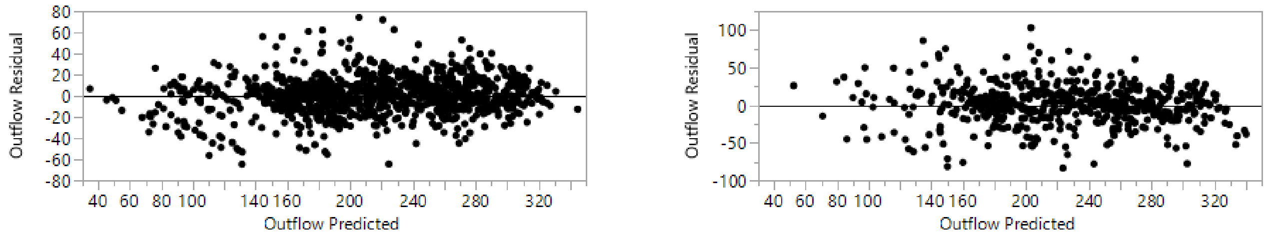


(a) (b)

Figure 3. (a) Predicted vs Observed Outflow (training); (b) Predicted vs Observed Outflow (validation)

4.3. Outflow Residuals

Outflow residuals versus predicted outflow plot was created to look at the distribution of residuals (Figure 4 (a),(b)). The residuals are also on the average equal to zero and there seems to be no systematic pattern to their distribution which means that the errors made by the model are purely random. This is a positive signal that the model is not overfitting and is making an unbiased prediction.



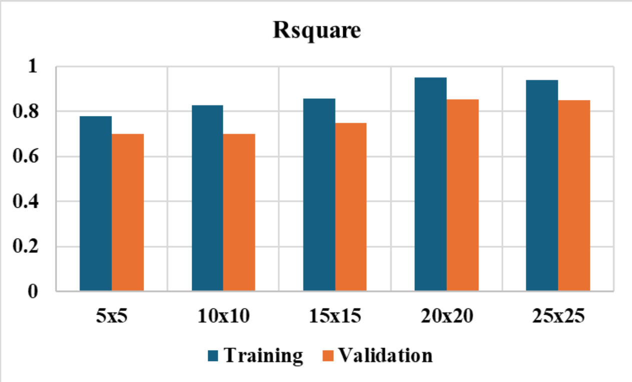
(a) (b)

Figure 4. (a) Outflow Residuals vs. Predicted Outflow (training); Outflow Residuals vs. Predicted Outflow (validation)

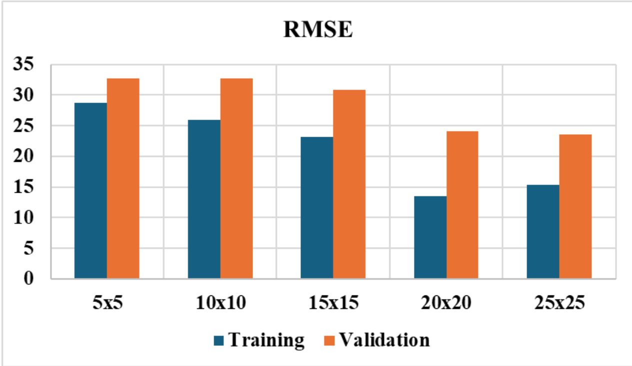
4.4. Model Performance and Evaluation

Through these models, the Artificial Neural Network (ANN) models were tested using the coefficient of determination (R^2), Root Mean Squared Error (RMSE) and Sum of Squared Errors (SSE) for the training and validation data set. From the results obtained, it can be concluded that the performance of the model increases with increase in the number of neurons in the hidden layers, and the 20 x 20 configuration offered the best results. Especially, it was observed that the 20x20 network had the highest training R^2 (0.9504), the lowest training RMSE (13.47) and SSE (221115.4) suggesting that this network provides a very good fit to the training data. Also, the 20x20 network provided the highest generalization capability as confirmed from the highest validation R^2 of 0.8543, the lowest validation RMSE of 24.06 and the lowest validation SSE of 352766.

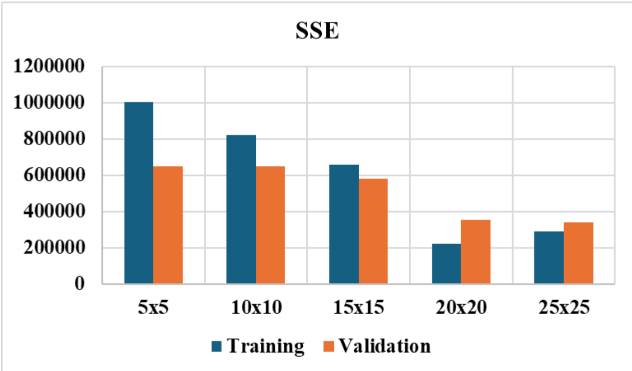
Subsequent computations also showed that the output performance were increased as the neurons increased from 5x5 to 20x20 but decreased in the 25x25 configuration. While the 25x25 model yielded a slightly smaller training RMSE of 15.3989 and SSE of the training data of 288582, the validation R^2 was insignificantly higher and the validation RMSE was insignificantly lower than the 20x20 model, indicating that overly complex models may overfit, to mean that while the model would approximate the training data very well, it poorly approximated the unseen Thus, the 20x20 is determined to be the most appropriate network model because of its capability of capturing the training data and at the same time possessing an excellent ability to generalise new patterns, and it is for this reason that this configuration has been adopted in this study. The comparison is presented in the Figure 5 (a), (b) and (c).



(a)



(b)



(c)

Figure 5. Comparison of model performance of various cases (a) R^2 (b) RMSE and (c) SSE

Table 1 also presents the ANN performance of all the cases using training and validation R^2, RMSE, and SSE..

Table 1. Comparison of R^2, RMSE, and SSE for Training and Validation Datasets Across Different ANN Configurations.

Neuron Configuration	Training R^2	Validation R^2	Training RMSE	Validation RMSE	Training SSE	Validation SSE
5x5	0.7788	0.7009	28.6985	32.6596	1002329	649589.02
10x10	0.8259	0.7009	25.9407	32.6596	818944.1	649589.02
15x15	0.8567	0.7482	23.232	30.7995	656844.8	577702.19
20x20	0.9504	0.8543	13.4792	24.0677	221115.4	352766.07
25x25	0.9378	0.8488	15.3989	23.5781	288582	338560.81

4.5. Interaction profiles, and model sensitivity

The sensitivity analysis through interaction profiles revealed that live storage is the most significant predictor of reservoir outflow due to its high main effect (84410) and comparatively low interaction with other variables (total effect: 9790). This underscores the importance of accurately measuring and incorporating live storage in reservoir management strategies. Moreover, inflow had moderate direct impact (0.097) and the highest overall impact (0.423) revealing it as another most significant and sensitive predictor of outflow. Rainfall exhibited minimal direct influence with main effect: 0.029) but significant interaction effects (total effect: 0.411) showing its moderate effects on outflow. Total seepage and maximum temperature had small main effects: 0.021 and 0.038, respectively, but their moderate total effects were 0.367 and 0.341, respectively, pointed to their indirect significance. In contrast, evaporation and relative humidity exhibited very small individual and combined impacts on outflow (main effect: 0 and 0.014; total effect: 29.03 and 0.129, respectively). Collectively, the results indicate that inflow and rainfall are the most influential predictors because of their interaction effects, and that evaporation and relative humidity have little influence on outflow. It is useful for fine-tuning current prediction algorithms and for improving the strategies being employed to manage the reservoir. The findings of this research are consistent with other studies that have used ANN for inflow and outflow prediction in reservoirs. Similarly, in the earlier research works, the authors [3] and [5] used similar success rate of ANN for the inflow forecasting in other reservoirs, thereby proving usefulness of this method in any climatic and hydrological environment.

Nevertheless, unlike previous studies that discussed the effectiveness of ANN in streamflow prediction under different situations [13], this study aims to investigate the ANN models for daily outflow prediction of Khanpur Reservoir in detail and discuss about its importance in real-time reservoir operation. These findings indicate that ANN learnt appropriately when the right input variables were used in its development and that optimum model tuning can make ANN a useful tool in reservoir management in semi-arid areas as supported by the level of accuracy in the discharge estimation during different climatic conditions.

5. Conclusion and Implications

The analysis of the results reported in this paper shows that the ANN model built in JMP can be used to predict the daily outflow of Khanpur Reservoir with sufficient level of accuracy and the results can be helpful in decision making regarding reservoir management. Due to the model's capability to consider non-linear interactions between hydrological and climatic factors, the model stands out as a valuable application in reservoir management strategies, especially in semi-arid areas such as Pakistan. The results also show that the appropriate choice of input parameters and further fine tuning of the model can further improve the performance of the model.

The interaction profiles and the contour plots obtained from the sensitivity analysis indicate that inflow and live storage are two of the most significant variables in the evaluation of the outflow while rainfall has the moderate significance level. The other variables has comparatively less significance in outflow prediction. Further studies should aim at extending the model by adding other independent variables such as soil moisture or seepage rates to increase the model performance and its reliability.

Acknowledgments: The authors are grateful to the Water and Power Development Authority (WAPDA) and the Punjab Irrigation Department for providing data from Khanpur Dam.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Al-Areeq, A. M., Al-Zahrani, M. A., & Sharif, H. O. (2021). The performance of physically based and conceptual hydrologic models: A case study for Makkah Watershed, Saudi Arabia. *Water, 13*(8), 1098. https://doi.org/10.3390/w13081098
2. ASCE Task Committee. (2000). Application of artificial neural networks in Hydrology 1: Preliminary concepts. *Journal of Hydrologic Engineering ASCE, 5*(2), 115–123.
3. Babaei, M., Moeini, R., & Ehsanzadeh, E. (2019). Artificial neural network and support vector machine models for inflow prediction of dam reservoir (case study: Zayandehroud dam reservoir). *Water Resources Management, 33*, 1–16. https://doi.org/10.1007/s11269-019-02252-5
4. Bozorg, H. O., Yari, P., Delpasand, M., & Chu, X. (2022). Reservoir operation under influence of the joint uncertainty of inflow and evaporation. *Environmental Development and Sustainability, 24*(2), 2914–2940.
5. Cheng, C., Niu, W., Feng, Z., Shen, J., & Chau, K. (2015). Daily reservoir runoff forecasting method using artificial neural network based on quantum-behaved particle swarm optimization. *Journal of Water, 7*, 4232–4246. https://doi.org/10.3390/w7084232
6. Chiamsathit, S., Adeloye, A. J., & Bankaru-Swamy, S. (2016). Inflow forecasting using artificial neural networks for reservoir operation. *International Association of Hydrological Sciences*.
7. Dawson, C. W., & Wilby, R. L. (2001). Hydrological modelling using artificial neural networks. *Progress in Physical Geography, 25*(1), 80–108.
8. Halwatura, D., & Najim, M. (2013). Application of the HEC-HMS model for runoff simulation in a tropical catchment. *Journal of Environmental Modelling & Software, 46*, 155–162.
9. HEC. (2021). U.S. Army Corps of Engineers, Hydrologic Engineering Center. HEC-HMS Hydrologic Modelling System, User's Manual, Version 4.8.0. Davis, CA. https://www.hec.usace.army.mil/software/HEC-HMS/documentation.aspx
10. Hu, S., & Shrestha, P. (2020). Examine the impact of land use and land cover changes on peak discharges of a watershed in the Midwestern United States using the HEC-HMS model. *Papers in Applied Geography*. https://doi.org/10.1080/23754931.2020.1732447
11. Jain, A., Maier, H. R., Dandy, G. C., & Sudheer, K. P. (2009). Rainfall-runoff modelling using neural networks: State of the art and future research needs. *Journal of Hydraulics Engineering, 15*(1), 52–74.
12. Jothiprakash, V., & Magar, R. B. (2012). Multi-time-step ahead daily and hourly intermittent reservoir inflow prediction by artificial intelligence techniques using lumped and distributed data. *Journal of Hydrology, 456*, 58–70. https://doi.org/10.1016/j.jhydrol.2012.04.045
13. Londhe, S., & Charhate, S. (2010). Comparison of data-driven modelling techniques for river flow forecasting. *Journal of Hydrological Sciences, 55*(7), 1163–1174.
14. Niu, W.-J., Feng, Z.-K., Feng, B.-F., Min, Y.-W., Cheng, C.-T., & Zhou, J.-Z. (2018). Comparison of multiple linear regression, artificial neural network, extreme learning machine, and support vector machine in deriving operation rule of hydropower reservoir. *Journal of Water, 11*, 88. https://doi.org/10.3390/w11010088
15. Vidyarthi, A., & Jain, S. (2020). A hybrid artificial intelligence model for streamflow forecasting. *Hydrology Research, 51*(4), 1029–1044. https://doi.org/10.2166/nh.2019.069
16. Niyazi, B. A., Hm Masoud, M., Admed, M., Basahi, J. M., & Rashed, M. A. (2020). Runoff assessment and modeling in arid regions by integration of watershed and hydrologic models with GIS techniques. *Journal of African Earth Sciences, 172*, 103966. https://doi.org/10.1016/j.jafrearsci.2020.103966
17. Okiria, E., Okazawa, H., Noda, K., Kobayashi, Y., & Suzuki, S. (2023). A comparative evaluation of lumped and semi-distributed conceptual hydrological models: Does model complexity enhance hydrograph prediction? *Hydrology, 9*(5), 89. https://doi.org/10.3390/hydrology9050089