Mr. Prabhu M, 2025, 13:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Leveraging Artificial Intelligence for Enhanced Stage-Discharge Curve Analysis in Hydrological Modelling

Mr. Prabhu M, Sanjeev Kumar. K, T.S. RameshBabu B.E
Assistant Director Gr-II, Central Water Commission, India

Abstract - This paper aims to assess the application of Artificial Intelligence (AI) in the development of stage-discharge relationships. Stage-Discharge Relationships often known as rating curves are a crucial tool for hydrological modelling and water resource management. This study uses AI, specifically Large Language Models (LLMs) in the development, calibration, and validation of stage-discharge curves. The methodology involves utilizing LLMs to generate python code. The hydrological data i.e stage and discharge from September 1971 to May 2021 (30 years) at Biligundulu gauging station on the Cauvery River was used to derive SD Curve. The developed equation was then applied to hydrological data from June 2021 to May 2023. The predicted discharge values were subsequently compared to the actual observed values. The outcomes are evaluated using R-squared, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), demonstrating a strong fit to the observed data (R^2 = 0.9995, RMSE = 17.29 m^3/s, MAE = 10.91 m^3/s). The process and results indicate that AI-driven approaches can offer a robust alternative to conventional methods. This study demonstrates AI can empower hydrologists without extensive coding expertise to conduct complex data analysis. By leveraging LLMs, complex hydrological models, data pre-processing, etc., can be automated, enabling more researchers and practitioners to conduct advanced analyses.

Keywords - Artificial Intelligence, Stage-Discharge Curve, Hydrological Modelling, Large Language Models, Water Resource Management, Predictive Accuracy, Nonlinear Regression.

I. INTRODUCTION

In hydrology, understanding the relationship between a river's stage (water level) and its discharge (flow rate) is fundamental for water resource management and flood forecasting (Maidment,1993). This relationship is commonly represented by a stage-discharge curve, often derived using power-law equations that empirically link stage and discharge (Herschy, 2009).

Q = C (h – h0)µ where C, h0 and µ are constants, and is a straight line on logQ, log (h – h0) axes (Fenton, J. D. , 2018).

Traditionally manual methods are used for developing these equations either empirically or automatically through spreadsheet software. However, these software tools lack in transparency on internal calculation and offer limited control over the fitting process. Further post processing capabilities for further analysis are limited, for comprehensive evaluation and refinement.

To address these limitations, this paper explores the application of Artificial Intelligence (AI), specifically Large Language Models (LLMs), to enhance the accuracy and efficiency of stage-discharge curve analysis (Goodfellow et al., 2016). By leveraging LLMs for data processing, model development, and validation, this study aims to demonstrate the potential of AI to transform hydrological data analysis and improve the reliability of discharge predictions. This approach not only offers a robust alternative to conventional methods but also paves the way for more informed decision-making in water resource management and risk mitigation.

The primary objectives of this study are: to develop a stage-discharge curve using AI algorithms based on measured stage data and observed discharge values; to assess the accuracy of predicted discharge values using statistical error metrics, such as Mean Absolute Error (MAE) and to explore the benefits and challenges of using AI in hydrological data analysis and water resource management.

II. STUDY AREA AND DATA DESCRIPTION

The study area is Biligundulu gauging station, located on the Cauvery River in Tamil Nadu, India, (12°10'56" N latitude, 77°43'26" E longitude). The Cauvery River is one of the major rivers of South India, originating in the Brahmagiri Hills of the Western Ghats and flowing through the states of Karnataka, Tamil Nadu, Kerala, and Puducherry before draining into the Bay of Bengal. The river basin is heavily populated and intensively cultivated, making water resource management a critical issue in the region.

This Hydro-Meterological station is operated by Central Water Commission, India. The station is strategically located on the border of States of Karnataka and Tamil Nadu [fig. 1]. It serves as a critical monitoring point for the Cauvery River, which makes a significant transition in its course both geographically, as it flows from the plateau to the plains, and geopolitically, as it lies at the inter-state boundary between Karnataka and Tamil Nadu. This location makes it a strategic site for water resource management and dispute resolution between the two states. Furthermore, the station plays a vital role as a base station for inflow forecasting for the Mettur Dam. The extensive dataset from this site, combined with dependable measurement protocols, makes it a reliable source for this study.

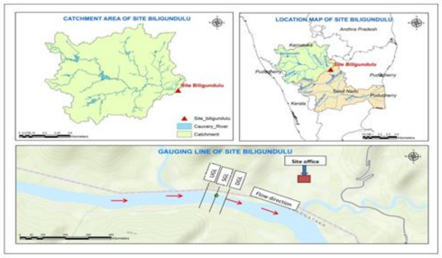


Figure 1 Map of Biligundulu Gauging Station, Tamilnadu, India

The daily discharge data observed at Biligundulu gauging station for the period September 3, 1971, to May 31, 2021 was used in this study for training and June 1, 2021 to May 31,2023 is used for testing. Discharge Observations are usually carried out at 08.00AM on all working days. The statistical parameters are given in table 1. In this table, count, mean, min,25%,50%,75%,max,std refers to number of observations, average, minimum, , the value below which 25% of the observations fall, 50% of the observations fall, 75% of the observations fall, maximum and standard deviation.

Table 1: Summary of Statistics of Training Data

	count	mean	min	25%	50%	75%	max	std
Observation Number	18296	1.00	1.00	1.00	1.00	1.00	2.00	0.02
stage	18296	3.52	2.40	2.98	3.25	3.80	10.65	0.82
Water Level	18296	258.52	257.40	257.98	258.25	258.80	265.65	0.82
discharge	18296	227.69	0.56	49.84	97.60	250.40	6691.70	377.18

Hh

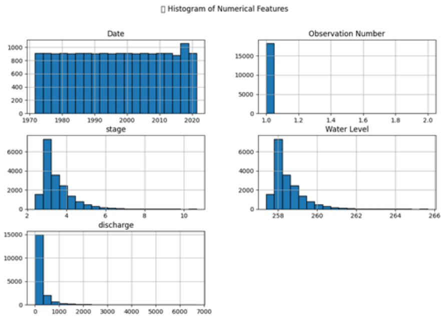


Figure 2:Histogram of Summary Statistics (Training data)-before prepocessing

The discharge values exhibit substantial variability, with a mean of 227.69 m^3/s, a standard deviation of 377.18 m^3/s, and a wide range from 0.56 to 6691.70 m^3/s. The high standard deviation relative to the mean suggests that the river experiences significant fluctuations in flow rate, potentially due to seasonal variations or extreme events. Histogram of the parameters of the discharge observation [fig. 2], such as an increased number of dates in the year 2017-18 indicates potential duplicate values.

III. METHODOLOGY

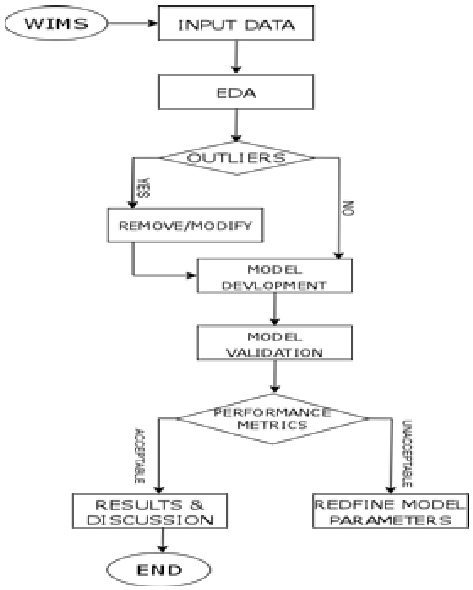
Artificial Intelligence (AI), specifically Large Language Models (LLMs), was used to develop and validate stage-discharge equation. The methodology involves following steps: data preprocessing, model development, model training, and model validation. Large Language Models (LLMs), were used to generate Python code workflow and for visualization.

The stage-discharge relationship was derived using power-law equation:
$Q = a * (H - H0)^b$,
where Q is the discharge (m^3/s), H is the gauge height (m), H0 is the datum correction (m), and a and b are empirical coefficients.

The parameters a, b, and H0 were estimated using nonlinear regression. The power equation model was created and validated using Python language within Google Colab environment. Colab is a cloud-based platform that is easily accessible and scientific computing libraries such as SciPy and Pandas, which are essential for data analysis and model fitting are pre-installed. This setup saves time and effort by removing the need to install and set up software. Colab can also be combined with Google Drive, for simpler storage and data management.

The Python code for implementing the project was generated using various large language models (LLM). LLM was utilized for the entire workflow, including data preprocessing, analysis, curve fitting and visualization. Prompts were provided to the LLM detailing the desired functionality—specifically, fitting a power law equation to stage-discharge data using non-linear least squares regression with scipy.optimize.curve_fit.

The LLM produced the Python code, including the function definition for the power law (rating_curve), the call to curve_fit with appropriate arguments (data, initial parameter estimates p0, and maximum iterations maxfev), and the extraction of the fitted parameters. The maximum number of iterations for the optimization algorithm (maxfev) was set to 10000. The generated code was executed and refined within the Google Colab environment, while data manipulation and visualization were carried out using the Pandas and Matplotlib libraries.



Results and Discussions

The power-law model, implemented with LLM-generated code, resulted in the following fitted rating curve equation for the Biligundulu station:

Q = 133.966 * (H - 2.400)1.840

The coefficient of determination (R^2) for the curve was calculated to be 0.9903, indicating a strong correlation and excellent goodness of fit [fig. 3].

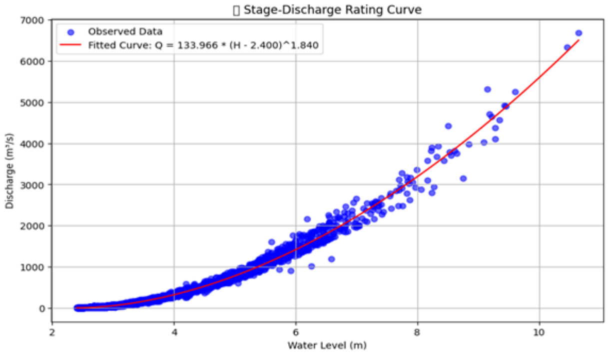


Figure 3: Stage Discharge Curve

Using this equation, data from June 1, 2021, to May 31, 2023 was validated and discharges were predicted. The summary statistics for both observed and predicted discharge values are presented in Table 2 . The 'count' was 730 for both observed and predicted values. The mean discharge values are 425.8776 and 420.30905 for observed and predicted, respectively. The standard deviation is 747.82 and 735.95 for observed and predicted, respectively showing a high variability in discharge values. The minimum observed discharge value is 6.509, and the maximum is 4879. The lower (25%), medium (50%), and higher(75%) quartiles for observed discharge are 56.775, 151.53, and 415.35, respectively.

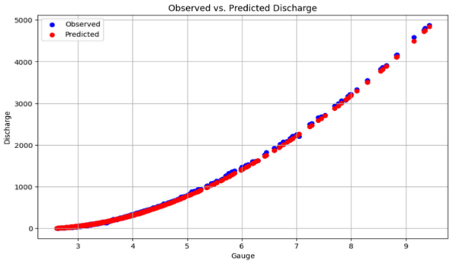


Figure 4: Observed vs Predicted Discharge

Table 2 : Summary of Statistics of Validated data

Statistic	Observed_Discharge	Predicted_Discharge
count	730	730
mean	425.8776	420.30905
std	747.826858	735.954365
min	6.509	7.583647
25%	56.775	60.639438
50%	151.53	162.326293
75%	415.35	399.138084
max	4879	4846.10019

Table 3: Goodness-of-Fit Statistics for Stage-Discharge Curve

Metric	Value
R-squared:	0.9995
RMSE:	17.29
MAE:	10.91
MAPE:	5.54%

The summary statistics for goodness of fit is presented in Table 3. The high R-squared value (0.9995) indicates that the fitted curve explains almost all of the variability in the observed discharge data. The low RMSE (17.29) and MAE (10.91) values suggest that the curve is, on average, quite close to the observed data points, both in terms of magnitude (MAE) and considering the influence of larger deviations (RMSE). The low MAPE (5.54%) indicates, on average, the model's predictions deviate from the actual (observed) discharge values by approximately 5.54%. Plot of observed vs predicted discharges is shown in fig.4

The residuals versus fitted discharge plot [fig. 5] reveals a non-random pattern, including a positive curvilinear trend, suggesting that the power-law model may not be fully capturing the underlying physics of the stage-discharge relationship.

Figure 5: Plot of Residuals vs Predicted Discharge

The plot also suggests that the model's predictions become less accurate at higher flows and some extreme values, particularly at higher discharges, which could be influencing the model fit. These observations imply that the model's predictions might not be as reliable during high flow events.

However, upon statistical examination, largest residual was 90 m3/s corresponding to a gauge of 9.15m and observed discharge of 4587 m3/s. On percentage terms it is less than 2%. The largest difference on positive side was 15% and negative side was -29%. The differences are showing in table 4 and table 5.

Table 4 : Top 5 Highest Percentage Differences on positive side

Gauge	Actual_Discharge	Predicted_Discharge	difference	% of Difference
2.95	52.651	44.59	8.06	15%
2.98	56.339	49.17	7.17	13%
3.06	70.42	62.37	8.05	11%
3.23	105.16	95.08	10.08	10%
3.22	102.83	92.98	9.85	10%

Table 5 : Top 5 Highest Percentage Differences on negative side

Gauge	Actual_Discharge	Predicted_Discharge	difference	% of Difference
2.9	28.96	37.42	-8.46	-29%
2.68	10.22	12.88	-2.66	-26%
2.68	10.27	12.88	-2.61	-25%
3.04	47.21	58.93	-11.72	-25%
3.52	132.24	165.03	-32.79	-25%

The histogram of residuals [Fig. 6] shows a desirable concentration of values close to zero, indicating that the fitted curve generally aligns well with the observed data. The spread of the residuals suggests that there is room for improvement in capturing all of the variations in the stage-discharge relationship

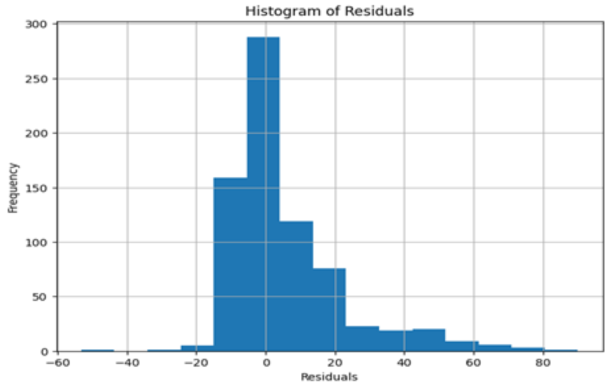


Figure 6: Histogram of residuals

IV. CONCLUSION

This study demonstrates that Artificial Intelligence (AI), specifically Large Language Models (LLMs) can be used to enhance the accuracy and efficiency of Stage Discharge Relationships. By leveraging LLMs, data processing, model development, validation, and visualization have been made robust and efficient, creating an alternative to traditional analysis without the need for complex coding software. This allows researchers to focus more on the study part.

The AI model developed in this study exhibits high accuracy, making it a valuable tool for predicting discharge values and improving our understanding of river flow dynamics. Moreover, the automation of code generation and data analysis tasks by LLMs opens up new possibilities for the wider adoption of AI in hydrological modeling.

Scope for improvement

This study can further be improved by seasonally decomposing the data for monsoon and non-monsoon periods. While the power-law model provided a reasonable fit to the observed data, future work could explore alternative functional forms, such as polynomial functions, logarithmic transformations, or more complex non-linear models, to potentially better capture the intricacies of the stage-discharge relationship.

Acknowledgment

The authors would like to express their sincere gratitude to the Central Water Commission (CWC), Cauvery & Southern Rivers Organisation, Coimbatore for providing the required data set. The authors would also like to thank Mr. TK Sivarajan, Chief Engineer (Retd) for his insightful comments and suggestions that contributed to the improvement of this work.

Prompts and Code used in the study

The prompts used for creating the code and resultant final python code used in the study are made publicly available. The link is https://github.com/prabhucwc/ai-sdcurve

REFERENCES

1. Atiaa, A. M. (2015). Modeling of stage-discharge relationship for Gharraf River,
2. southern Iraq by using data driven techniques: A case study. In E.W. Publications, Water Utility Journal (Vol. 9, pp. 31–46) [Journal-article]. https://www.ewra.net/wuj/pdf/WUJ_2015_09_04.pdf
3. Bisht, D. C. S., 1, Raju, M. M., 2, & Joshi, M. C., 3. (n.d.). ANN BASED RIVER STAGE – DISCHARGE MODELLING FOR GODAVARI RIVER, INDIA. In Computer Modelling and New Technologies: Vol. Vol.14 (Issue No.3, pp. 48–62). https://www.tsi.lv/sites/default/files/editor/science/Research_journals/Computer/2010/V3/14_3-6_bisht.pdf
4. Cheng, Z., Lee, K., Kim, D., Muste, M., Vidmar, P., & Hulme, J. (2018). Experimental evidence on the performance of rating curves for continuous discharge estimation in complex flow situations. Journal of Hydrology, 568, 959–971. https://doi.org/10.1016/j.jhydrol.2018.11.021
5. Chow, V. T. (1959). Open Channel Hydraulics. McGraw-Hill.
6. Fenton, J. D. (2018). On the generation of stream rating curves. Journal of Hydrology, 564, 748–757. https://doi.org/10.1016/j.jhydrol.2018.07.025
7. Guven, A., Aytek, A., & Azamathulla, H. M. (2012). A practical approach to formulate stage–discharge relationship in natural rivers. Neural Computing and Applications, 23(3–4), 873–880. https://doi.org/10.1007/s00521-012-1011-5
8. Guerrero, J., Westerberg, I. K., Halldin, S., Xu, C., & Lundin, L. (2012). Temporal variability in stage–discharge relationships. Journal of Hydrology, 446–447, 90–102. https://doi.org/10.1016/j.jhydrol.2012.04.031
9. Hydrology Project. (1999). How to establish stage discharge rating curve. In https://www.mahahp.gov.in/images/29HOWTOESTABLISHSTAGEDISCHARGERATINGCURVE.pd
10. Jiang, Z., Wang, H., & Song, W. (2013). Discharge estimation based on machine learning. DOAJ (DOAJ: Directory of Open Access Journals). https://doi.org/10.3882/j.issn.1674-2370.2013.02.003
11. Kumar, M., Kumari, A., Kushwaha, D. P., Kumar, P., Malik, A., Ali, R., & Kuriqi, A. (2020). Estimation of daily Stage–Discharge relationship by using Data-Driven techniques of a perennial river, India. Sustainability, 12(19), 7877. https://doi.org/10.3390/su12197877
12. Mistry, K., Patel, C., Gundalia, Dr. M., & Chhotubhai Gopalbhai Patel Institute of Technology, Surat, India. (2023). DETERMINATION OF STAGE-DISCHARGE RELATIONSHIP OF DAMAN GANGA RIVER, WESTERN INDIA. In Journal of Emerging Technologies and Innovative Research (Vol. 10, Issue 6). https://www.jetir.org/papers/JETIR2306673.pdf
13. Modelling of Stage-Discharge relationship. (2018). https://www.indianjournals.com/ijor.aspx?target=ijor:ndcwwc&volume=7&issue=1&article=002
14. Muzzammil, M., Alam, J., & Zakwan, M. (Eds.). (2015). N Optimization Technique for Estimation Of Rating Curve Parameters [National Symposium Of Hydrology -2015]. Indian Association Of Hydrologists (Iah), Roorkee. https://www.researchgate.net/publication/305767208_AN_OPTIMIZATION_TECHNIQUE_FOR_ESTIMATION_OF_RATING_CURVE_PARAMETERS
15. Roy, S., & Mistri, B. (2013). Estimation of peak flood discharge for an ungauged river: A case study of the Kunur River, West Bengal. Geography Journal, 2013, 1–11. https://doi.org/10.1155/2013/214140
16. Sefe, F. T. K. (1996). A study of the stage-discharge relationship of the Okavaiigo River at Mohembo, Botswana. Hydrological Sciences Journal, 41(1), 97–116. https://doi.org/10.1080/02626669609491481
17. Srinivas.M.L (1999)., Understanding Stage-Discharge Relations. Hydrology Project- Training

of Trainers in Hydrometry. https://nhp.mowr.gov.in/Docs/HP-2/Manuals/UnderstandingStage-DischargeRelat.pdf

18. Vias, R. D., Hrafnkelsson, B., Hodson, T., Rögnvaldsson, S., Jansson, A. Ö., & Gardarsson, S. M. (2024). Evaluating the applicability of the generalized power-law rating curve model. Journal of Hydrology, 132537. https://doi.org/10.1016/j.jhydrol.2024.132537
19. Westerberg, I., Guerrero, J., Seibert, J., Beven, K. J., & Halldin, S. (2010). Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras. Hydrological Processes, 25(4), 603–613. https://doi.org/10.1002/hyp.7848
20. WMO (2010). Manual on Stream Gauging (Volume I & II). World Meteorological Organization, Report No. 1044