International Journal on Engineering Technology and Infrastructure Development (InJET-InDev)
www.lec.edu.np/journal
Volume 2, Issue no. 2, Page 113-123
Received Date: 11/24/2025
Revised Date: 5/27/2026
Accepted Date: 5/31/2026
ISSN: 3021-999X (Print)
https://doi.org/10.3126/injet-indev.v2i2.95706

Physics-Informed Data Augmentation for Sediment Concentration Prediction in Data-Scarce Himalayan Rivers

Usan Adhikari[1*], Mukesh Raj Kafle[2], Sushan Adhikari[3]

[1]*IOE Tribhuvan University, Pulchowk, Lalitpur Nepal, adhikariusan@gmail.com*
[2]*IOE Tribhuvan University, Pulchowk, Lalitpur Nepal, mkafle@pcampus.edu.np*
[3]*Kathmandu University, Dhulikhel, Kavre Nepal, sushan.adhikari2060@gmail.com*

Abstract

The correct prediction of sediment concentration has prime significance in the development of sustainable hydropower in the Himalayan region because of the large sediment loads being hazardous to the turbines as well as the reservoirs' longevity. This study tackles the challenges of the scarcity of data and the complexity of physics through the organized comparisons of the conventional statistical and advanced physics-based data augmentation techniques which can be used again for the sediment prediction driven by the ML approach in the Himalayan region characterized by the scarcity of data. This study pursued two interrelated objectives: first, organized comparisons of conventional and advanced data augmentation schemes; and second, the development of new physics-informed schemes for sediment transport to create a reproducible analytics framework suitable for data-scarce Himalayan watersheds. In this study, the effectiveness of ten data augmentation techniques: five classical statistical ones (forward-backward fill, linear interpolation, seasonal mean approach, simple rating curve models, and ensemble averaging) and five advanced models founded on physics (seasonal stochastic rating curve models, k-nearest neighbor discharge analogs, STL decomposition models, physics-based constraints models, and weighted ensemble) was investigated using the same number of observed monthly sediment data points. Conservative pre-processing of the data resulted in the preservation of about 99.8% of the data points via the consensus approach of three methods of outlier detection. The advanced models based on physics were greatly superior to the classical statistical models for the augmentation of sediment concentrations regarding the enhancement of R^2 value performance metric (by 5.5%), the Root Mean Squared Error (RMSE—by 24.3%), and the Mean Absolute Error (by 47.6%) all tested through rigorous 5-fold cross-validation. This study makes various contributions. Firstly, this study can be classified as research in the field of hydrology due to its subject of addressing various data scarcity challenges in this discipline. Additionally, this study lays the groundwork for further machine learning analysis by making sure that missing sediment data can be imputed effectively with minimal imputation errors.

Keywords: Sediment Transport Prediction, Data Augmentation, Physics Informed Modeling, Himalayan Hydrology, Hydropower Engineering, Rating Curves

1. Introduction

Nepal has considerable hydropower resources due to its geographical location in the Himalayas, estimated at approximately 83,000 MW of which about 42,000 MW can be economically harnessed (Shrestha, 1966). However, only a limited capacity of about 3,602 MW has been harnessed until 2025 (Ministry of Finance, Government of Nepal, 2025). Most of the already operational schemes that were built are run-of-river schemes that rely on the discharge offered in the rivers during particular seasons of the year. This has resulted in the development of storage-based hydropower schemes like the Budhigandaki Hydroelectric Project of 1,200 MW capacity because of their ability to provide constant power production (BGHEPDC, 2020). One of the challenges to the sustainability of the above storage schemes is sedimentation. The Himalayan rivers are known to be sediment-laden owing to their steep slopes, vulnerability to earthquakes, and the concept of monsoon climate in Nepal, which results in the deposition of sediment inside the reservoirs (Chalise & Khanal, 1997; Mahmood, 1987; White, 2001). Precise sediment estimation helps in the management of the above risks. However, the conventional models fail in Nepal because of the unavailability of reliable data and the non-linear nature of sediment transport (Horowitz, 2003; Kişi, 2010). In this regard, the possibility of using machine learning algorithms has been found valuable in the estimation of sediment loads due to the ability of the algorithms to

* *Corresponding author*

manage challenging data patterns and combine various data sources such as remote sensing (Hassan et al., 2022). This paper will emphasize the various data collection, processing, visualization, and augmentation methodologies that can be used to improve the quality of the data sets used in developing machine learning algorithms to address this problem in the Himalayan regions.

In the sedimentation-prone Himalayan environment, the reservoirs face the dilemma of the rapid depletion of their active storage capacity due to the continuous accumulation of sedimentations (Dahal et al., 2024). The sediment-laden water causes constant abrasive wear on turbine blades and runner surfaces, progressively degrading efficiency and increasing maintenance costs (Pradhan, 2004). The case of the Kulekhani Reservoir illustrates this impact due to the continuous depletion of its storage capacity since 1982. The conventional method of sediment measurement and forecasting utilizes large amounts of high-quality hydrological and sediment data. This type of data can be inadequate and unreliable in Nepal and other less developed regions (Magaju et al., 2020). The measurement of sediment itself constitutes a labor-intensive and expensive procedure involving various methods of measurement through depth integration and point sampling of sediments during the monsoon season combined with the measurement of concentration and grain size through laboratory work (Nepal sediment studies). Himalayan rivers are among the most dynamic systems globally, owing to their steep topography, fragile geology, and the significant influence of monsoon patterns, which collectively result in exceedingly high sediment loads and highly variable flow conditions (Sinha et al., 2019; Winrock International, 2019). In major basins like the Koshi, Gandaki, and Karnali, their shape changes from narrow V-shaped valleys to wide braided channels. This is because of steep slopes, weak formations like the Siwalik Hills, and heavy rain during the monsoon season, which cause erosion, landslides, and large sediment pulses (Ghimire et al., 2013; Marc et al., 2019; Nayak, 1993).

Sediment flows as bed load, suspended load, or wash load. Suspended load is the hardest for hydropower systems to deal with, and its movement is controlled by rules like the Shields criterion and stream power theory (Shields, 1936; Bagnold, 1966; Yang, 1973; Yang et al., 1996). The Himalayas are even more complicated because they have strong seasonality, with 70–85% of annual rainfall falling in just four months (Bookhagen & Burbank, 2006). They also have glacier-derived fine sediments at high elevations (Immerzeel et al., 2010), and landslide-driven sediment surges happen often because of active tectonics (Ghosh et al., 2020). These conditions also make hydrological data likely to have gaps. Simple imputation methods like linear interpolation, forward/backward filling, seasonal means, and rating-curve approaches don't work well because they assume linear behavior and miss extreme variability (De Boor, 1978; Box & Jenkins, 1976; Horowitz, 2003; Little &

Rubin, 2019). More advanced technologies like STL decomposition, analogue/k-NN matching, and physics-informed neural networks can help, but they still have problems in the Himalayas, where there isn't much data (Cleveland et al., 1990). Best practice suggests combining physical reasoning, hydrological-regime stratification, and cross-validation. However, there is still no clear comparison between traditional and physics-guided augmentation methods for Himalayan sediment data. This study fills that gap by looking at how well they work and how they affect the reliability of sediment prediction.

In this research paper, we target the challenges of data scarcity through the improvement of the basic stages of sediment modeling: data gathering, processing, mapping, and executing conventional as well as advanced data augmentation techniques. The above objectives will help the research lay the cornerstone for a possible model of sediment prediction through the usage of machine learning algorithms without necessarily embarking on the specifics of the model and its output. This research paper aims at developing and applying the concept of data augmentation based on the physics of sediment transport to mitigate the scarcity of data in the Himalayan river basin context, especially at the location of the Budhigandaki Hydropower Project. The prime aim of this paper is to improve the available data about sediment and hydrology in the region using conventional and advanced methods of data augmentation.

2. Materials and Methods

2.1 Study Area

The Budhigandaki River Basin is located in a transnational region bordered by the Tibetan Plateau to the north, the Marsyangdi River Basin on the west, and the Trisuli River Basin on the east and south. The drainage area at the proposed dam site covers about 5,005 km², with around 30% in Tibet. Elevations range from approximately

310 m to 8,000 m, with diverse climatic zones from subtropical lowlands to polar high altitudes. The basin experiences a strong southwest monsoon providing 80–90% of annual precipitation, significantly controlling hydrological and sediment dynamics.

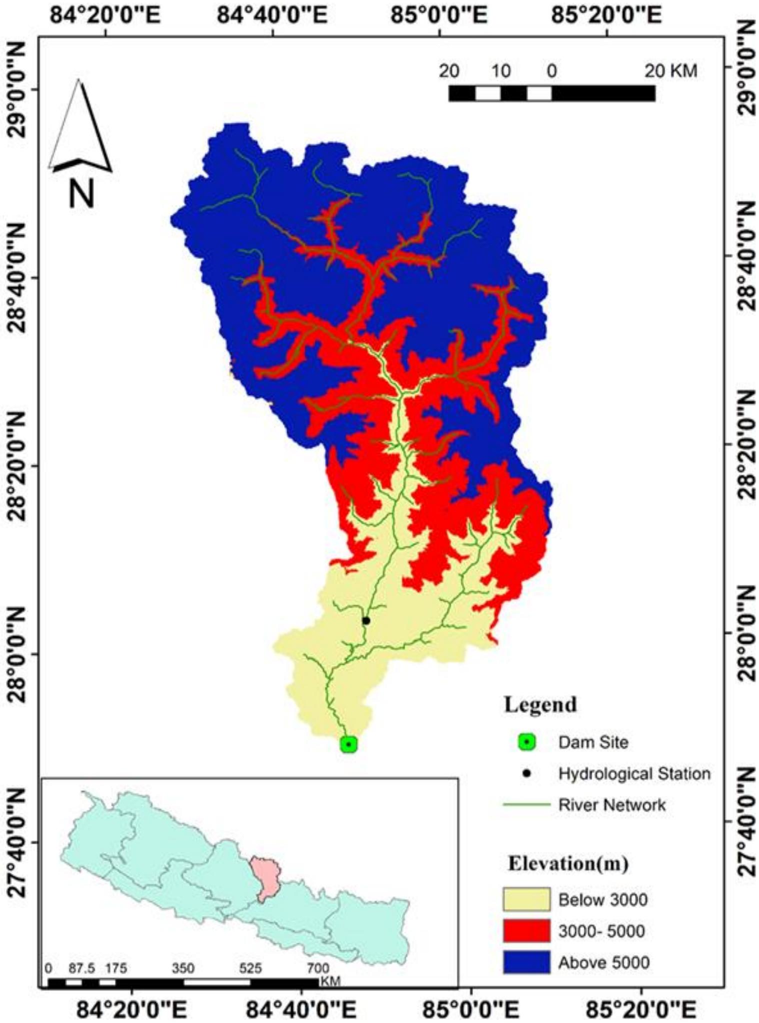

Figure 1. Watershed Map of Budhigandaki River

The Nepal Electricity Authority (NEA) is developing a reservoir-style hydropower project called the Budhi Gandaki Hydropower Project. When finished, this 1,200 MW hydropower plant will be the biggest in Nepal. Its purpose is to control river flow to produce electricity, which will lessen reliance on imported petroleum-based energy and create a large number of jobs. A concrete double-curvature arch dam that is 263 meters high and a reservoir with active storage of about 4,467 million cubic meters are two of the dam's main features. A crucial hydrological record for the basin is provided by discharge data at Arughat Station (1964–2014), with interpolation filling in brief data gaps. Jade Consult's sedimentological studies are the source of suspended sediment

concentration data for the same time period. One of the longest continuous sediment records for a Himalayan river, the dataset retains 560 valid monthly observations with few missing values after stringent quality control. In addition to hydrological and sediment data for integrated analysis, climate variables from NASA POWER provide additional predictors for the 1981–2019 period, such as temperature, precipitation, humidity, wind speed, and surface pressure. In addition to a statistically significant long-term increasing trend in sediment levels, which is probably caused by shifting land use and climatic factors, the sediment concentration time series shows clear seasonality with monsoon peaks and dry-season lows. Between the monsoon and dry seasons, the mean concentration of sediment is 31:1, and the relationships between sediment and discharge show different seasonal transport regimes. In this complicated tropical basin, correlations show that precipitation is a better predictor of sediment than discharge, highlighting the effects of both hydraulic capacity and sediment supply.

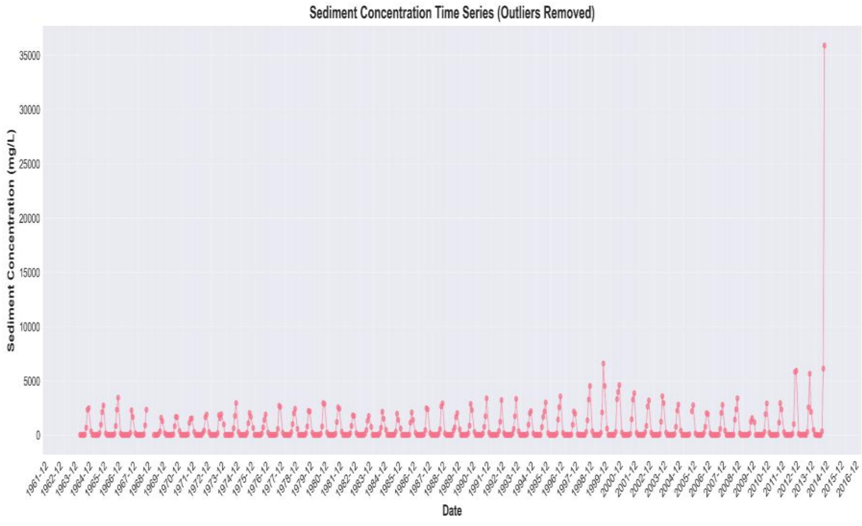

Figure 2. Sediment Concentration Time Series

Significant month-to-month variability spanning 2.2 orders of magnitude is revealed by quantitative characterization of seasonal sediment concentration patterns (Figure 2). While July and August monsoon peaks show explosive increases (mean 2,600–3,100 mg/L) accompanied by extreme variability (standard deviation 1,000–4,800 mg/L, reflecting inter-event and inter-annual differences), January through April show remarkably stable low concentrations (mean 18–19 mg/L, standard deviation 2–4 mg/L).

Following outlier removal and quality control, the final analysis dataset comprises:

- Time period: January 1964–December 2014 (51 years)
- 560 monthly averages (560 of 612 potential = 91.5% completeness) are valid observations.
- Eight sediment concentrations (1.4%) and five discharge values (0.9%) are missing.
- Outliers eliminated: 0.2% of observations

2.2 Methodology

In order to separate the effects of data augmentation techniques from model selection confounders, this study uses a controlled comparison methodology. Robust evaluation of augmentation impact is made possible by applying identical downstream processing pipelines to datasets augmented by both sophisticated physics-informed and conventional statistical techniques. Three sequential steps make up the methodology:

- Stage 1: Data Collection and Quality Assessment in which Sediment, discharge, meteorological, and temporal data from 1964 to 2014 are integrated with conservative outlier removal maintaining 99.8% of observations.
- Stage 2: Dual-Track Data Augmentation uses five sophisticated physics-based methods in addition to five conventional imputation techniques. It is validated through 5-fold cross-validation and demonstrates improved prediction metrics with physics-informed methods.
- Stages 3 and 4: Additionally, features can be engineered and applied in rigorous comparative analysis and machine learning model applications, but these are outside the purview of this paper.

Analyses are performed in Python 3.9 with widely adopted scientific libraries—pandas, numpy, scikit-learn, scipy, statsmodels, matplotlib, and seaborn—to ensure reproducibility. Fixed random seeds guarantee consistent results across platforms, with full codebases documented for transparency. To fully evaluate model quality, several complementary metrics are used, including R^2, RMSE, MAE, Nash-Sutcliffe Efficiency, and Kling-Gupta

Efficiency. In addition to better capturing a variety of aspects of prediction performance pertinent to sediment concentration modelling, this multi-metric approach reduces the biases inherent in single metrics. Extreme outliers are found using a cautious three-method consensus process that strikes a balance between eliminating measurement errors and preserving physically plausible monsoon peaks:

- Extreme worth Monthly averages are capped at 50,000 mg/L.
- Season-Aware Interquartile Range maintains acceptable seasonal variability by employing distinct thresholds for the monsoon and dry seasons.
- Physical consistency is checked using sediment-discharge rating curve residuals, noting deviations greater than 500%.

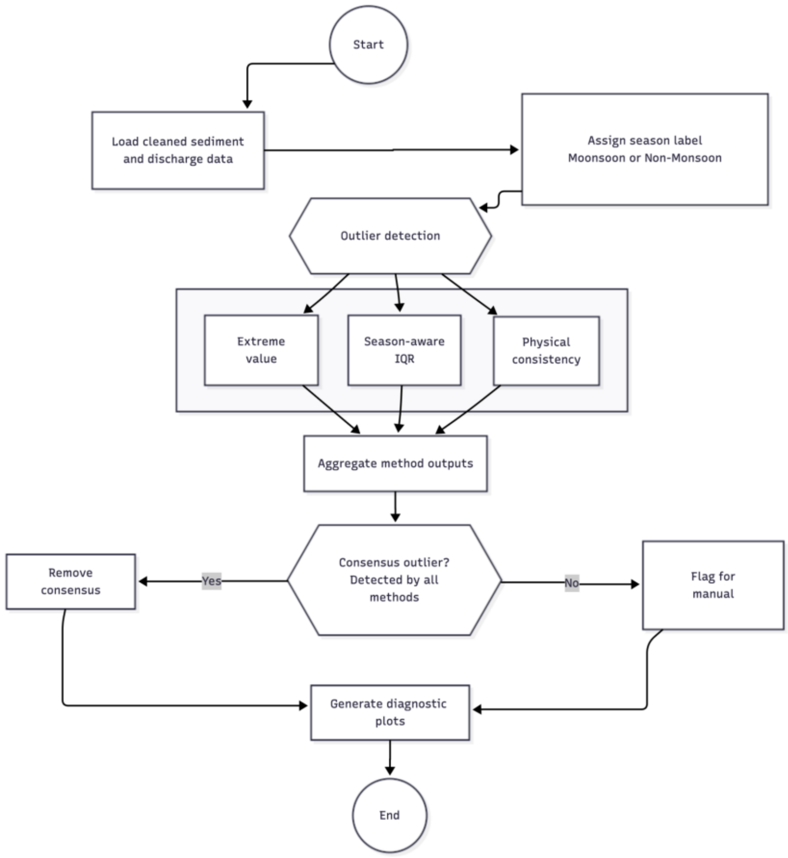

Figure 3. Traditional Augmentation Method Flowchart

Sediment transport Physics is not specifically incorporated into statistical imputation techniques used in traditional data augmentation methods. These act as starting points for evaluating the effectiveness of physics-based methods. Eight (1.4%) of the 560 monthly sediment concentration records (January 1964–December 2014) were missing. Careful imputation is essential to maintain time series continuity, optimize data utilization, and prevent bias if missingness is non-random, even with low missingness.

To ensure impartial performance estimation, validation employed 5-fold cross-validation with artificially masked values, comparing imputed and actual data using R^2, RMSE, and MAE metrics averaged over holdout sets.

Traditional Data Augmentation Methods: Forward-Backward Fill offers precise preservation of valid values and computational simplicity by propagating adjacent observations forward or backward. However, it may increase measurement errors, ignore temporal trends and discharge data, and produce artificial plateaus. In linear interpolation method, straight lines connecting neighboring valid points create smoother series and marginally better metrics. Nevertheless, the approach ignores discharge influence and imposes unrealistic linear temporal trends. In seasonal mean imputation method, long-term monthly averages are used to replace missing values, successfully capturing strong monsoonal seasonality (roughly 30× annual variation). Loss of inter-annual variability and disregard for discharge fluctuations are among the limitations. The other method is sediment rating curve imputation in which sediment concentrations were estimated from discharge measurements using the power-law sediment-discharge relationship fit to all data $C_{predicted} = aQ^b$. Using the traditional ensemble method, metrics showing a modest incremental gain are balanced, and variance is reduced by an equal-weighted average of the four previous imputations. Although seasonal and discharge inclusion gradually improves traditional methods, basic limitations still exist: lack of physics constraints, insufficiency of a single rating curve, incapacity to model hysteresis or uncertainties, inadequate handling of extreme events, and lack of antecedent moisture memory. These encourage the next methodological step to investigate sophisticated

physics-informed augmentations. In order to produce realistic sediment data for data-poor Himalayan rivers, this paper uses a number of physics-informed augmentation techniques that explicitly incorporate sediment transport dynamics. By incorporating uncertainty quantification and physical understanding, the methods outperform conventional statistical approaches.

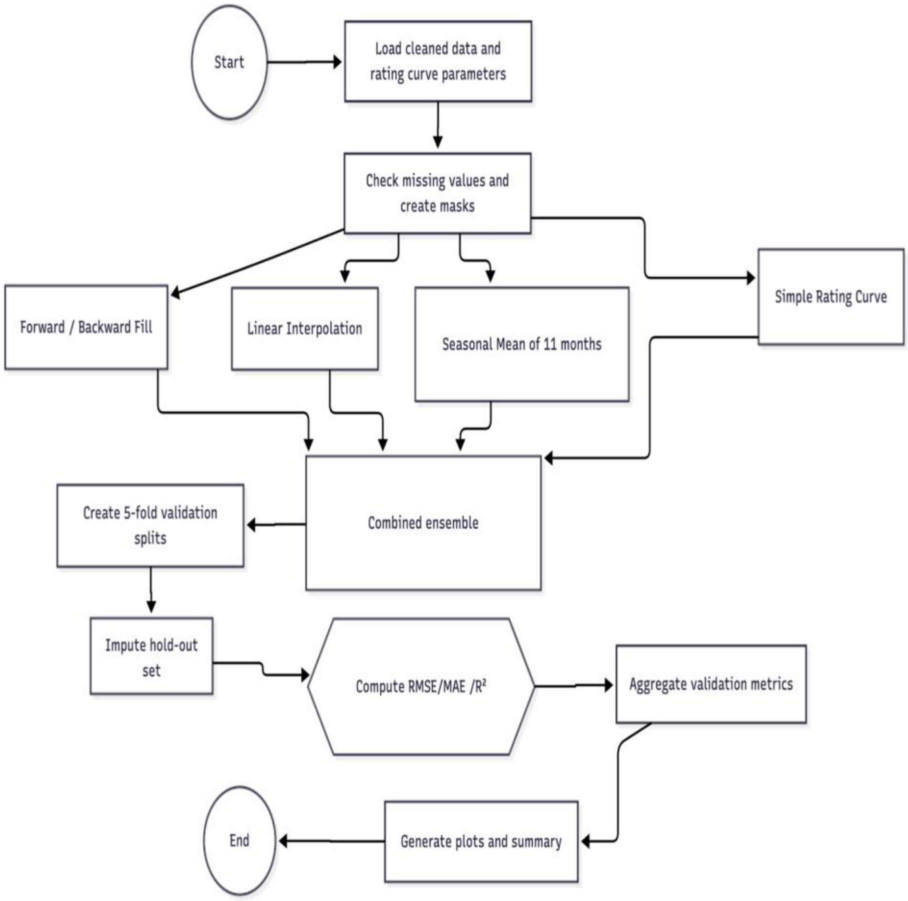

Figure 4. Augmentation Method Flowchart

Advanced Data Augmentation Methods: For the monsoon and dry seasons, the Seasonal Deterministic Rating Curves approach uses distinct sediment-discharge power-law relationships that represent essentially different transport regimes (transport-limited vs. supply-limited). Although this approach is deterministic and does not represent uncertainty, it does capture seasonal variability. Another method, the Stochastic Rating Curves work by including an uncertainty term that models natural variability caused by storm intensity variation, channel storage dynamics, and episodic landslides, this extends deterministic curves. This method produces ensemble data with confidence bounds that are essential for engineering design. Next method, the k- Nearest Neighbor analog method work by identifying past months with comparable river discharge, analogues in discharge space use hydraulic similarity rather than temporal proximity to infer sediment concentrations. This nonparametric data-driven approach manages nonlinearities and yields physically significant imputations. The STL Decomposition method decomposes the time series into trend, seasonal, and residual components, imputing missing values by adding estimated trend and seasonality while excluding unpredictable residual noise.

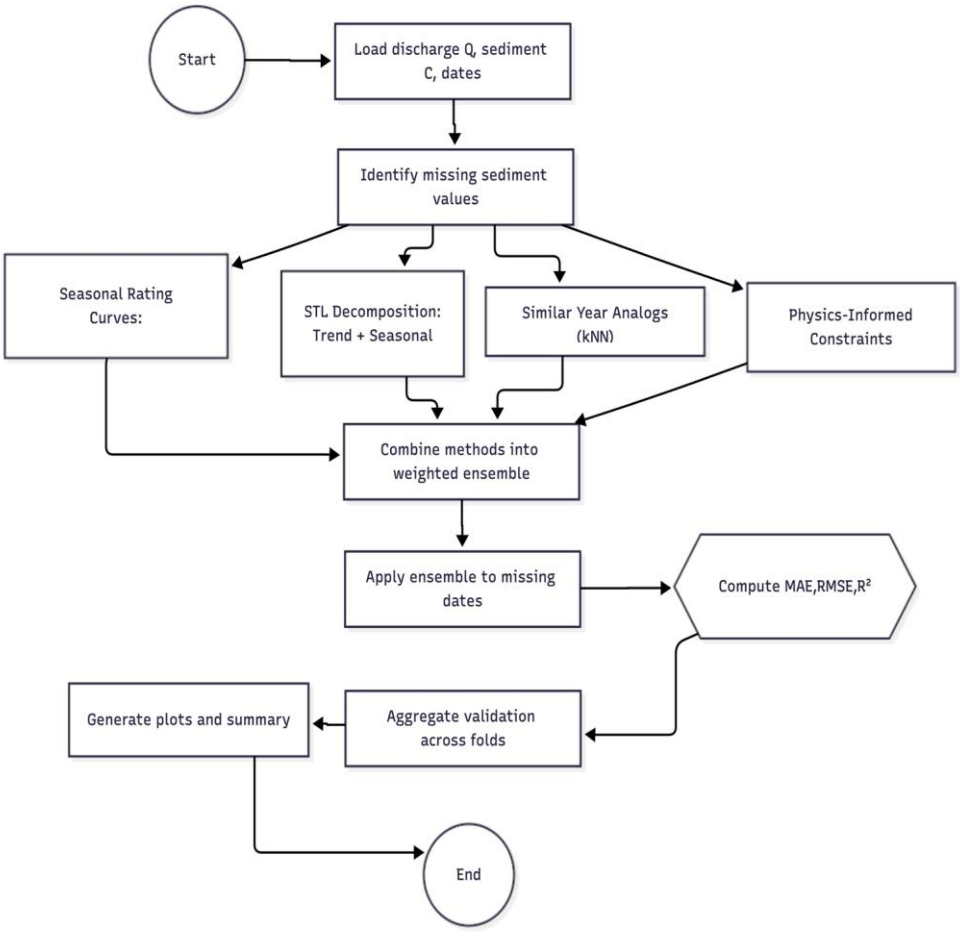

Figure 5. Advanced Augmentation Method Flowchart

This approach adapts to variable seasonality and long-term changes but is less effective around abrupt transitions. The advanced weighted ensemble method combines the above methods according to their predictive reliability and physical relevance, followed by the application of physics constraints to finalize augmented data. The cross-validation framework procedure offers an objective assessment of how well augmentation techniques reconstruct missing sediment data, providing crucial information about their suitability prior to additional modelling.

10% of sediment observations should be randomly masked to mimic missing data.

- For the remaining 90%, apply either the traditional or advanced augmentation method.
- Compare the original masked values with the imputed values.
- Calculate performance metrics (R^2, RMSE, MAE) based on the augmentation outcomes.
- To create a five-fold cross-validation, repeat the procedure with five distinct random masks.
- To get reliable estimates of augmentation effectiveness, average metrics across folds.

3. Results and Discussion

3.1 Outlier Detection Results

The three-method consensus retained 99.8% of all observations (559 of 560 monthly records). The single removed point (August 2014) was flagged by all three detection criteria: it exceeded the 50,000 mg/L absolute threshold, fell outside the season-aware IQR bounds for monsoon months, and deviated from the rating curve residual by more than 500%. No other observations triggered unanimous consensus removal, confirming that the conservative threshold design preserved physically plausible extreme events while eliminating only confirmed measurement anomalies.

3.2 Traditional Statistical Methods Performance

Table 1: Performance of Traditional Statistical Augmentation Methods

Method	R^2 Mean ± SD	RMSE (mg/L)	MAE (mg/L)
Forward-Backward Fill	0.517 ± 0.091	787 ± 126	490 ± 97
Linear Interpolation	0.575 ± 0.106	737 ± 138	459 ± 99
Seasonal Mean	0.720 ± 0.066	591 ± 63	292 ± 28
Simple Rating Curve	0.779 ± 0.067	534 ± 137	**229 ± 34**
Traditional Ensemble	**0.804 ± 0.067**	**507 ± 146**	288 ± 70
Best	0.804	507	288

Forward-backward fill performs poorly due to lack of trend and physics consideration, causing unrealistic constant values.

- Despite losing year-to-year variability, the seasonal mean significantly improves accuracy in capturing monsoon-driven seasonality.
- By utilizing discharge data, rating curves that incorporate sediment transport physics enhance predictions even more.
- By combining strengths, an ensemble of methods reduces the risk of overfitting and produces the best overall performance.

While seasonality and physical insights are gradually added to traditional methods, they are still constrained by assumptions like single rating curves, the absence of uncertainty modelling, and the incapacity to capture extreme events or cumulative antecedent effects. For dependable data augmentation in sparse-data environments, these gaps encourage the adoption of sophisticated physics-based augmentations that more accurately reflect the dynamics and uncertainty of Himalayan sediments.

3.3 Advanced Physics-Based Methods Performance

- By weighting historical sediment data according to hydraulic similarity, k-NN analogues outperform single methods in terms of accuracy and computational efficiency.
- Due to seasonal subset constraints, seasonal deterministic rating curves show somewhat greater variability despite offering reliable predictions based on sediment transport theory.

Table 2: Advanced Augmentation Models Performance

Method	R^2 Mean ± SD	RMSE Mean ± SD (mg/L)	MAE (mg/L)
Seasonal Deterministic	0.833 ± 0.132	408 ± 226	177 ± 57
Seasonal Stochastic	0.771 ± 0.142	479 ± 226	208 ± 65
k-NN Analogs	0.842 ± 0.134	394 ± 229	**134 ± 47**
STL Decomposition	0.771	—	—
Advanced Ensemble	**0.848±0.135**	**384±232**	151±49
Best	0.848	384	151

RMSE and MAE are not reported for STL Decomposition as it is a descriptive smoothing algorithm rather than a predictive model; it does not generate point predictions and therefore cannot be evaluated with error metrics. R^2 is reported based on the reconstructed series against observed values.

- In order to balance somewhat lower point accuracy with useful confidence estimates for engineering design, seasonal stochastic curves introduce uncertainty quantification.
- While STL decomposition is good at capturing seasonality and long-term trends, it has trouble with sudden monsoon transitions.
- The advanced ensemble integrates data-driven flexibility and a theoretical foundation to produce the best overall performance by combining methods with suitable weights and constrained by physics-based rules.

These advanced augmentation methods collectively improve upon traditional statistical approaches by embedding sediment transport physics, accounting for uncertainty, and leveraging hydraulic similarity, resulting in more realistic and reliable augmented datasets suitable for subsequent predictive modelling in Himalayan sediment contexts.

3.4 Traditional Statistical vs Advanced Physics-Based Augmentation Methods Performance

Table 3: Comparison of Traditional and Advanced Augmentation Methods Performance

Metric	**Traditional**	**Advanced**	**Difference**
R^2 (Mean ± SD)	0.804 ± 0.067	**0.848 ± 0.135**	+5.5%
RMSE (mg/L)	507 ± 146	**384 ± 232**	-24.3%
MAE (mg/L)	288 ± 70	**151 ± 49**	-47.6%
CV Stability	0.067	**0.135**	+101.5%

Advanced physics-based augmentation techniques perform better than traditional statistical methods because they incorporate basic principles of sediment transport and explicitly model natural variability. These techniques produce realistic, physically plausible synthetic data that more accurately depict intricate seasonal and event-driven sediment dynamics in Himalayan rivers by taking into account physical limitations like supply limits and transport capacities and using hydraulic similarity for analogue selection. The slight increase in cross-validation metric variance reflects this enhanced model flexibility, capturing uncertainty essential for robust engineering design.

The markedly higher cross-validation variance in advanced methods (SD of R^2 = 0.135 vs. 0.067 for traditional) warrants careful interpretation. This increased variability is not indicative of instability but rather reflects the greater sensitivity of physics-informed methods to the heterogeneous nature of Himalayan sediment regimes — particularly the sharp contrast between monsoon and dry-season dynamics. In engineering applications, this trade-off is acceptable: the advanced ensemble's substantially lower mean errors (RMSE: 384 vs. 507 mg/L; MAE: 151 vs. 288 mg/L) are of greater practical consequence for reservoir design than fold-to-fold variance. Nevertheless, practitioners should be aware that performance on individual folds covering monsoon-transition months may show elevated error, and ensemble weights should be recalibrated when applying this framework to other basins.

These results align with findings from physics-informed data augmentation studies in analogous data-scarce basins. Studies on Himalayan and trans-Himalayan river systems (e.g., Kosi, Gandaki) have similarly demonstrated that single rating curves systematically underestimate peak monsoon sediment concentrations due to hysteresis and supply-limited transport (Sinha et al., 2019; Horowitz, 2003). The superior performance of k-NN analogs in this study is consistent with nonparametric approaches reported in other high-variability fluvial systems, where hydraulic similarity outperforms temporal proximity for imputation (Hassan et al., 2022). However, direct comparison remains limited by the scarcity of published augmentation benchmarks specifically for Himalayan sediment datasets — a gap this study begins to address.

3.5 Limitations

Several limitations of this study should be acknowledged. First, the analysis is conducted on a single basin (Budhigandaki), and transferability to other Himalayan river systems with different geological settings or monsoon intensities has not been validated. Second, ensemble weights for the advanced methods were derived empirically from the available 51-year record and may require recalibration for shorter or sparser datasets. Third, the cross-validation framework masks 10% of observations randomly, which may not fully simulate

real-world missingness patterns that tend to cluster during peak monsoon months — precisely when accurate data are most critical. Finally, the evaluation is limited to imputation performance; the downstream impact of augmented data on ML model predictive accuracy remains a topic for follow-on work.

4. Conclusion

Future ramifications include enhanced predictive accuracy of sediment concentration models that support ecosystem conservation and hydropower sediment management. By optimizing the usefulness of scarce data, such physics-informed augmentation can lessen reliance on costly, sparse observational campaigns and open the door to scalable, transferable models that can be applied to a variety of riverine systems under climatic and human pressures. This research framework thus contributes a critical methodological advance toward sustainable river basin management in sediment-challenged Himalayan environments.

In order to improve sediment concentration datasets in Himalayan river basins with limited data, this study highlights the importance of advanced, physics-based data augmentation techniques. By combining hydrological and sediment transport physics, these improved datasets greatly outperform traditional statistical imputations in capturing intrinsic seasonal dynamics, transport capacities, and natural uncertainties. The enhanced data quality will provide a stronger foundation for future hydropower management and sediment prediction models. The demonstrated improvements in predictive fidelity and uncertainty estimation not only solve long-standing problems in Himalayan sediment modelling, but they also establish a methodological framework that can be applied to other complex, data-constrained fluvial systems. Plans for augmentation are expected to improve in the future by taking climate change projections and sediment source variability into account.

In the end, this research contributes to a more comprehensive geo-scientific understanding of Himalayan sediment dynamics while bridging important data gaps and developing useful tools that support sustainable hydropower development.

References

Bagnold, R. A. (1966). *An approach to the sediment transport problem from general physics.* US government printing office.

Bookhagen, B., & Burbank, D. W. (2006). Topography, relief, and TRMM-derived rainfall variations along the Himalaya. *Geophysical Research Letters*, *33*(8), 2006GL026037. https://doi.org/10.1029/2006GL026037

Box, G., & Jenkins, G. M. (1976). *Analysis: Forecasting and Control.*

Chalise, S. R., & Khanal, N. R. (1997). Erosion processes and their implications in sustainable management of watersheds in Nepal Himalayas. *IAHS Publication*, *246*, 325–334.

Cleveland, R. B., Cleveland, W. S., & McRae, J. E. (1990). STL: A seasonal-trend decomposition. *Journal of Official Statistics*, *6*(1), 3–73.

Dahal, V., Kunwar, S., Bhandari, S., Chaudhary, S., Gautam, S., Bhatt, N., & Regmi, R. K. (2024). Analyzing sedimentation patterns in the Naumure Multipurpose Project (NMP) reservoir using 1D HEC-RAS modeling. *Scientific Reports*, *14*(1), 22134. https://doi.org/10.1038/s41598-024-73883-x

De Boor, C. (1978). *A practical guide to splines* (Vol. 27). Springer.

Ghimire, S., Higaki, D., & Bhattarai, T. (2013). Estimation of Soil Erosion Rates and Eroded Sediment in a Degraded Catchment of the Siwalik Hills, Nepal. *Land*, *2*(3), 370–391. https://doi.org/10.3390/land2030370

Ghosh, S., Bose, S., Mandal, N., & Laik, A. (2020). Mid-crustal ramping of the Main Himalayan Thrust in Nepal to Bhutan Himalaya: New insights from analogue and numerical experiments. *Tectonophysics*, *782–783*, 228425. https://doi.org/10.1016/j.tecto.2020.228425

Hassan, S., Shaukat, N., Ahmad, A., Abid, M., Hashmi, A., Shahid, M. L. U. R., Rajabi, Z., & Tariq, M. A. U. R. (2022). Prediction of the Amount of Sediment Deposition in Tarbela Reservoir Using Machine Learning Approaches. *Water*, *14*(19), 3098. https://doi.org/10.3390/w14193098

Horowitz, A. J. (2003). An evaluation of sediment rating curves for estimating suspended sediment concentrations for subsequent flux calculations. *Hydrological Processes*, *17*(17), 3387–3409. https://doi.org/10.1002/hyp.1299

Immerzeel, W. W., Van Beek, L. P. H., & Bierkens, M. F. P. (2010). Climate Change Will Affect the Asian Water Towers. *Science*, *328*(5984), 1382–1385. https://doi.org/10.1126/science.1183188

Kişi, Ö. (2010). River suspended sediment concentration modeling using a neural differential evolution approach. *Journal of Hydrology*, *389*(1–2), 227–235. https://doi.org/10.1016/j.jhydrol.2010.06.003

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data.* John Wiley & Sons.

Magaju, N., Sharma, D., Poudel, A., & Aryal, M. (2020). Identification of economically feasible run-of-river hydropower sites: A case study of West Rapti River. *Journal of Engineering and Environmental Studies.*, *4*(2), 45–56.

Mahmood, K. (1987). *Reservoir sedimentation: Impact, extent, and mitigation (World Bank Technical Paper No. 71)*. World Bank.

Marc, O., Behling, R., Andermann, C., Turowski, J. M., Illien, L., Roessner, S., & Hovius, N. (2019). Long-term erosion of the Nepal Himalayas by bedrock landsliding: The role of monsoons, earthquakes and giant landslides. *Earth Surface Dynamics*, *7*(1), 107–128. https://doi.org/10.5194/esurf-7-107-2019

Ministry of Finance, Government of Nepal. (2025). *Economic survey 2024/25*. Government of Nepal.

Nayak, J. K. (1993). Sediment management of the Kosi River basin, Nepal. *IAHS Publication*, *217*, 583–586.

Pradhan, B. (2004). Sediment erosion in low specific speed Francis turbines: A case study on effects and causes. *Proceedings of the International Conference on Hydraulic Engineering*, 134–148.

Shields, A. (1936). *Application of similarity principles and turbulence research to bed-load movement.*

Shrestha, J. N. (1966). *Water resources of Nepal and their utilization.* Ministry of Water Resources, Government of Nepal.

Sinha, R., Gupta, A., Mishra, K., Tripathi, S., Nepal, S., Wahid, S. M., & Swarnkar, S. (2019). Basin-scale hydrology and sediment dynamics of the Kosi River in the Himalayan foreland. *Journal of Hydrology*, *570*, 156–166. https://doi.org/10.1016/j.jhydrol.2018.12.051

White, W. R. (2001). *Evacuation of sediments from reservoirs*. Thomas Telford Publishing.

Winrock International. (2019). *Nepal water resources profile overview.* Winrock International.

Yang, C. T. (1973). Incipient Motion and Sediment Transport. *Journal of the Hydraulics Division*, *99*(10), 1679–1704. https://doi.org/10.1061/JYCEAJ.0003766

Yang, C. T., Molinas, A., & Wu, B. (1996). Sediment Transport in the Yellow River. *Journal of Hydraulic Engineering*, *122*(5), 237–244. https://doi.org/10.1061/(ASCE)0733-9429(1996)122:5(237)