International Journal on Engineering Technology (InJET)

www.kec.edu.np/kec-journal/
Volume 3, issue no. 1, Dec 2025, Page 240-270
Received Date: Sep 02, 2025
Accepted Date: Sep 16, 2025
ISSN: 3021-940X (Print)
DOI: https://doi.org/10.3126/injet.v3i1.87131

Comparison of Sediment Rating Curve Developed by Using KNN, SVM and Random Forest Method for Gandaki River Basin, Nepal

Kamal Katwal[1*], Sudeep Thapa[2], Dipendra Bajracharya[3], Gaurab Ranjit[4], Isha Karna[5], Bishal Pudasaini[6],

[1]*Senior Lecturer, National College of Engineering, Lalitpur, Nepal, kamal@nce.edu.np*
[2]*Junior Associate Professor, Kantipur Engineering College, Lalitpur, Nepal, sudeepthapa@kec.edu.np*
[3] *Graduate Student, Kantipur Engineering College,Lalitpur, Nepal, deepeshbajracharya1@gmail.com*
[4] *Graduate Student, Kantipur Engineering College,Lalitpur, Nepal, gaurabranjit@gmail.com*
[5] *Graduate Student, Kantipur Engineering College,Lalitpur, Nepal, isha.karna1276@gmail.com*
[6] *Graduate Student, Kantipur Engineering College,Lalitpur, Nepal, bishalpudasaini9869@gmail.com*

Abstract

Accurate estimation of suspended sediment load (SSL) is crucial for effective water resource management in Nepal's monsoon-influenced Himalayan River systems, where traditional sediment rating curves (SRCs) often fail to capture complex, non-linear discharge-sediment relationships. This study developed season-specific sediment rating curves for Station 430.5 on the Seti-Gandaki River using three machine learning algorithms: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest (RF) to address the limitations of conventional power-law approaches across Nepal's four distinct hydrological seasons. Daily discharge and suspended sediment concentration data from 2004-2017 were analyzed to characterize seasonal variability, with monsoon periods exhibiting peak flows exceeding 1,400 m^3/s and sediment concentrations reaching over 20,000 ppm, contrasting sharply with winter baseflow conditions below 35 m^3/s and sediment concentrations under 250 ppm. Model performance was evaluated using coefficient of determination (R^2) and Mean Absolute Percentage Error (MAPE), revealing optimal algorithm selection varied significantly across seasons: KNN demonstrated superior accuracy for monsoon ($R^2 = 0.22$) and post-monsoon ($R^2 = 0.97$) periods, while SVM showed better performance during winter ($R^2 = 0.81$) conditions, though all models struggled with pre-monsoon transitional dynamics ($R^2 < 0.25$). The resulting power-law equations—$S = 0.85 \times Q^{1.81}$ (pre-monsoon), $S = 333.66 \times Q^{0.31}$ (monsoon), $S = 26.55 \times Q^{0.62}$ (post-monsoon), and $S = 17.61 \times Q^{0.72}$ (winter) captured distinct seasonal sediment transport mechanisms, with monsoon conditions presenting the greatest modeling challenges (MAPE = 90.67%) due to complex hysteresis effects, variable sediment availability, and extreme non-linearity that conventional machine learning approaches cannot fully represent. While machine learning models demonstrated substantial improvements over traditional SRCs, particularly during stable post-monsoon and winter periods, the persistent high prediction errors during monsoon conditions highlight the need for advanced deep learning architectures capable of capturing temporal dependencies and multi-factorial sediment transport controls. These findings provide essential foundations for improving sediment load estimation in data-scarce Himalayan basins and inform sustainable water resource management, hydropower planning, and climate adaptation strategies in rapidly changing mountain river systems.

Keywords: sediment rating curve, machine learning, suspended sediment load, seasonal hydrology, Himalayan rivers, discharge-sediment relationship, KNN regression, support vector machine, random forest, Gandaki basin

1. Introduction

Sediment transport modelling in river systems has emerged as a critical component of water resource management and environmental monitoring, particularly in complex mountainous watersheds where traditional methodologies often fail to capture the intricate relationships between hydrological variables and sediment dynamics (Carbonneau, Lane and Bergeron, 2006). The Gandaki River Basin, situated in the Central Himalayas of Nepal, represents one of the world's most dynamic sediment transport systems, where extreme

topographic relief, monsoon-driven precipitation patterns, and glacial processes create highly variable sediment loading conditions (Shrestha and Aryal, 2011) (Baniya *et al.*, 2019). Traditional sediment rating curves (SRCs), which establish empirical relationships between water discharge and suspended sediment concentration, have long been the standard approach for sediment load estimation; however, these conventional methods frequently exhibit significant limitations in accurately predicting sediment transport, particularly during extreme flow events and seasonal variations (Asselman, 2000).

The deficiencies of traditional sediment rating curves are well-documented, with studies demonstrating systematic underestimation of suspended sediment loads during high discharge periods and overestimation during low flow conditions (Boukhrissa *et al.*, 2013) (Oeurng, Sauvage and Sánchez-Pérez, 2011). This phenomenon is particularly pronounced in Himalayan river systems, where complex hydrological processes, including snowmelt contributions, glacial lake outburst floods, and monsoon-induced runoff, create non-linear relationships between discharge and sediment concentration that cannot be adequately captured by simple power-law functions (Baniya *et al.*, 2019). Furthermore, the temporal variability in sediment transport mechanisms, characterized by distinct seasonal patterns and hysteresis effects, necessitates the development of more sophisticated modelling approaches that can accommodate the dynamic nature of sediment-discharge relationships (Morin *et al.*, 2018a).

Machine learning approaches have demonstrated superior performance in addressing the limitations of traditional sediment transport models by capturing complex, non-linear patterns in hydrological data that conventional statistical methods cannot adequately represent (Tfwala and Wang, 2016) (Banadkooki *et al.*, 2020) (Abid Almubaidin *et al.*, 2023). Advanced algorithms, including artificial neural networks (ANNs), support vector machines (SVMs), random forest (RF), and deep learning architectures such as Long Short-Term Memory (LSTM) networks, have shown remarkable capability in modelling intricate temporal dependencies and capturing seasonal variations in sediment transport processes (Khosravi *et al.*, 2022) (Katwal *et al.*, 2025) (Van, Le and Chin, 2023) (Jamal Chachan and Sulaiman Bahnam, 2022) (Nda *et al.*, 2023). Recent studies have reported significant improvements in prediction accuracy, with machine learning models achieving coefficient of determination values exceeding 0.9 compared to traditional methods, while effectively reducing root mean square errors by up to 40% (Bajracbajharya *et al.*, 2025) (Salarijazi *et al.*, 2024). The integration of seasonal adjustment techniques with Bayesian optimization has further enhanced model performance, demonstrating the potential for developing robust, season-specific sediment rating curves that can accommodate the temporal complexities inherent in Himalayan River systems (Chhetri *et al.*, 2016) (Bajracharya *et al.*, 2025).

This research addresses the critical need for developing accurate, season-specific sediment rating curves in the Gandaki River Basin through the application of advanced machine learning methodologies that can effectively capture the complex temporal and spatial variability in sediment transport processes characteristic of high-mountain river systems. The study aims to overcome the limitations of traditional approaches by leveraging the superior pattern recognition capabilities of machine learning algorithms to develop robust predictive models that can inform sustainable water resource management and environmental planning in this ecologically and economically important watershed.

2. Material and Methodology

The primary data source for this study consists of suspended sediment concentration and discharge measurements collected by the Department of Hydrology and Meteorology (DHM), Nepal at monitoring station 430.5 located on the Seti-Gandaki River within the Gandaki River Basin, spanning the period from 2004 to 2017. A comprehensive analysis of the dataset revealed a significant disparity in data availability, with discharge measurements providing complete temporal coverage (100%) throughout the study period, while suspended sediment concentration data exhibited substantial gaps with only 33.79% temporal coverage over the same timeframe. This pronounced data scarcity in sediment measurements represents a critical constraint for effective water resource planning and sediment transport assessment in the basin, necessitating

the development of robust predictive methodologies that can leverage the consistently available discharge data to estimate suspended sediment concentrations. To address this fundamental limitation and enhance the accuracy of sediment load predictions for improved river basin management, this research was specifically designed to develop and validate machine learning-based predictive models that can reliably estimate suspended sediment concentration using the comprehensive discharge dataset as the primary input variable (Figure 1).

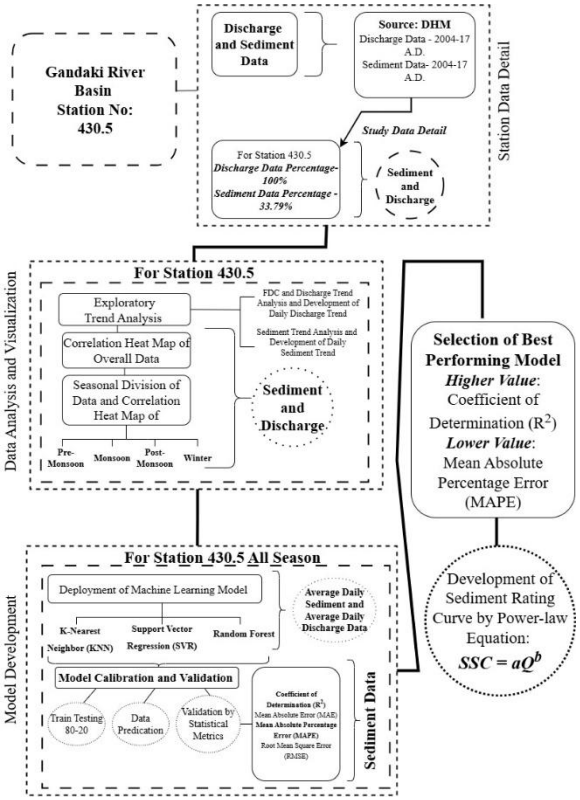

Figure 1: Flowchart

2.1 Study Area "Gandaki River Basin"

The Gandaki River Basin (Figure 2), situated in the Central Himalayas of Nepal, originates from high-altitude glacial and snowmelt sources on the Tibetan Plateau and descends through some of the world's deepest gorges before joining the Ganges in northern India (Pant *et al.*, 2025). Encompassing a catchment area of approximately 43 300 km², the basin exhibits a dramatic altitudinal gradient from over 7 000 masl. in the Annapurna and Dhaulagiri ranges to under 300 masl. in the Terai lowlands resulting in highly variable hydrodynamic and sediment transport regimes throughout its course as shown in figure 3.

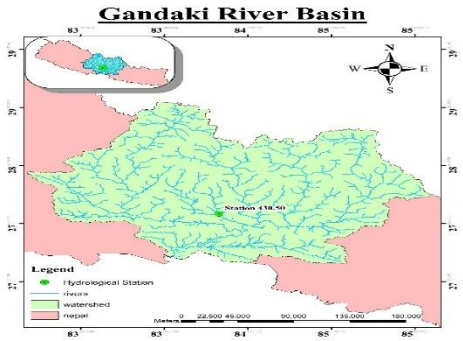

Figure 2: Gandaki River Basin

The basin's complex drainage network is characterized by numerous tributaries including the Seti Gandaki, Myagdi, and Marshyangdi rivers that drain steep mountain slopes and converge in deep valleys before entering the floodplains (Oli and Zomer, 2010). Geoprocessing analyses of ALOS PRISM-derived DEMs reveal a highly dendritic drainage pattern, with channel density and bifurcation ratios reflecting intense tectonic uplift and prolonged fluvial incision (Pant *et al.*, 2018). These geomorphic controls govern sediment yield and distribution, creating distinct sediment transport zones along the river's longitudinal profile.

Climatically, the basin experiences a monsoonal regime, with over 80 percent of annual precipitation occurring between June and September, while winter precipitation is dominated by snowfall in the higher catchments (Sharma *et al.*, 2020). Snowmelt contributions are critical to sustaining baseflow during the pre-monsoon dry season, with SRM-based studies indicating that snowmelt accounts for 20–30 percent of total annual discharge in the Budhi Gandaki sub-basin (Sharma *et al.*, 2020). This pronounced seasonality drives strong temporal variability in sediment transport, with peak suspended sediment concentrations occurring during the monsoon and meltwater pulses.

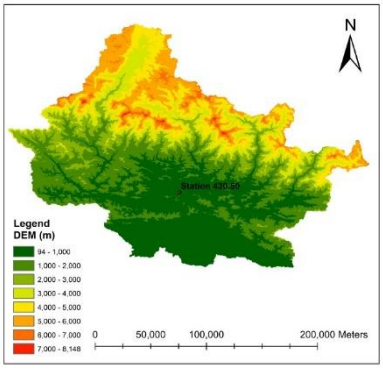

Figure 3: DEM of Study Area

Beyond its hydrological importance, the Gandaki Basin supports rich biodiversity and provides vital ecosystem services and livelihoods for over four million inhabitants. The Seti Gandaki tributary alone harbors more than 30 fish species across five orders, whose distribution correlates strongly with physicochemical gradients such as discharge, dissolved oxygen, and substrate composition (Prasad *et al.*, 2020). Agricultural irrigation, hydropower development, and sediment management are therefore closely intertwined with the basin's ecological integrity and socio-economic well-being.

2.2 Discharge and Sediment Data Processing

The analysis of discharge and sediment trends at Station No. 430.5 in the Gandaki River Basin begins with the assessment of the frequency distribution of daily discharge, which reveals clear seasonal flow patterns characterized by high peak discharges during the monsoon and markedly low flows in the dry season. Parallel examination of temporal variations in suspended sediment concentration illustrates how sediment transport fluctuates in response to these discharge dynamics. Because discharge records spanned the entire period from 2004 to 2017 while sediment measurements were sporadic and covered only about one-third of days (33.79%), average daily sediment values were derived from available observations through statistical aggregation and interpolation. Both discharge and sediment series were then converted into synchronized daily averages, creating a consistent dataset that accurately reflects their relationship under varying hydrological conditions.

To ensure data quality and consistency, outliers were identified and removed using the 1.5 × interquartile range rule, whereby any value beyond 1.5 × the IQR above Q3 or below Q1 was excluded from further analysis (Aggarwal, 2017). The cleaned dataset was subsequently split into training and testing subsets 80 percent for model calibration and 20 percent for validation while maintaining the chronological order of observations (Hyndman and Koehler, 2006). This rigorous preprocessing and partitioning protocol provided a robust foundation for developing and evaluating predictive models of suspended sediment concentration based on daily discharge.

2.3 Seasonal Variation and its importance in SRC

Nepal's river systems are governed by a distinct four-season hydrological cycle; pre-monsoon (April–June), monsoon (June–October), post-monsoon (October–December) and winter (December–March)—which, combined with steep Himalayan topography, results in strong seasonal shifts in both discharge and sediment transport (Chhetri *et al.*, 2016) (WECS, 2011). During the monsoon, intense rainfall and accelerated glacial melt produce peak flows and mobilize over 80 % of the annual water discharge and sediment load, whereas moderate snowmelt-driven flows characterize the pre-monsoon period and low-flow conditions dominate the post-monsoon and winter months (WECS, 2011) (Andermann *et al.*, 2012)(Chhetri *et al.*, 2016). These seasonal dynamics not only alter sediment magnitude but also change the discharge–suspended sediment concentration relationship, with stronger hysteresis during the monsoon complicating traditional regression-based sediment rating curves (Chhetri *et al.*, 2016) (Morin *et al.*, 2018b) . Incorporating seasonality into sediment rating curve development is therefore crucial for enhancing predictive accuracy, particularly in data-limited Himalayan basins where reliable sediment estimates inform water resource management, hydropower design and flood risk mitigation.

2.4 Machine Learning in Sediment Rating Curve (SRC)

Sediment rating curves (SRCs) traditionally rely on empirical, often linear regressions between river discharge and suspended sediment concentration to estimate suspended sediment load (SSL). While these conventional methods are straightforward, they frequently fail to represent the complex, non-linear, and time-dependent interactions that drive sediment transport, especially under highly variable flow regimes such as monsoon floods. Phenomena like hysteresis, variable sediment availability, and antecedent moisture conditions introduce significant deviations from the simple discharge–sediment relationship, causing traditional approaches to underestimate SSL during peak flows and complicate sediment budgeting, reservoir siltation assessments, and infrastructure design (Ponce, 2023) (Walling, Webb and Webb, 1988). To address these shortcomings and improve predictive accuracy across all flow conditions, there is a clear need to adopt machine learning (ML) techniques that can capture non-linear patterns, temporal dependencies, and multivariate influences inherent in hydrological and sediment transport processes.

Among the spectrum of available ML algorithms, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Random Forest (RF) emerge as particularly suitable for SRC development due to their complementary strengths. KNN offers a non-parametric, instance-based framework capable of detecting local discharge–sediment patterns with minimal assumptions, making it valuable for basins exhibiting stable flow dynamics (Ezzaouini *et al.*, 2022) . SVM employs kernel functions to project input variables into higher-dimensional spaces, effectively modeling complex, non-linear relationships influenced by factors such as rainfall and land use (Azamathulla *et al.*, 2010). RF, as an ensemble of decision trees, excels in reducing prediction variance and overfitting while uncovering intricate interactions among predictors, as evidenced by Nash–Sutcliffe efficiency values between 0.47 and 0.80 in SSL prediction studies (Ezzaouini *et al.*, 2022). The demonstrated performance of these algorithms in diverse river systems justifies their selection for developing a robust, season-specific SRC at Station No. 430.5 in the Gandaki River Basin, where accurate SSL estimation is critical for sustainable water resource management.

2.5 Model Parameter Tuning

a) K-Nearest Neighbors (KNN)

For the K-Nearest Neighbors regression model, systematic hyperparameter optimization was conducted to achieve optimal performance, with particular emphasis on determining the ideal number of neighbors (k), which fundamentally controls the bias-variance trade-off in the model. A comprehensive grid search was implemented across k values ranging from 1 to 31, utilizing 5-fold cross-validation to ensure robust performance assessment (Hastie, Tibshirani and Friedman, 2009) . This cross-validation approach partitioned the dataset into five equal subsets, iteratively training on four folds while validating on the remaining fold, providing a comprehensive evaluation across different data configurations. The K-Neighbors Regressor from scikit-learn library was employed for implementation, with each k value evaluated using multiple regression metrics including coefficient of determination (R^2), mean squared error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). Cross-validation scores were averaged across all folds to obtain stable performance estimates, with the optimal k value selected based on maximizing R^2 while minimizing error metrics, ensuring improved accuracy and generalization for suspended sediment load prediction.

b) Support Vector Machine (SVM)

Support Vector Machine regression was implemented to model the complex non-linear relationship between river discharge and suspended sediment concentration, leveraging kernel functions to project input data into higher-dimensional feature spaces where linear separation becomes feasible. The SVM's effectiveness in capturing intricate sediment transport dynamics is highly dependent on careful selection of key hyperparameters, particularly the kernel function type and its associated parameters. Three primary kernel functions were evaluated: linear (requiring only the regularization parameter C), radial basis function (RBF) requiring both C and gamma parameters, and polynomial kernels utilizing C and degree parameters. Grid search cross-validation was employed to systematically explore the hyperparameter space, with the RBF kernel typically demonstrating superior performance in handling non-linear hydrological relationships due to its ability to capture complex patterns in high-dimensional spaces. The optimal hyperparameter configuration was determined through 5-fold cross-validation, ensuring robust model selection and minimizing overfitting while maximizing predictive accuracy for sediment load estimation.

Table 1 Parameter Description Table SVM

Hyperparameter	Description	Initialization
Kernel	Specifies the kernel function used to map data to a higher dimensional space (e.g. linear, radial basis function, poly, sigmoid, etc.).	Radial Basis Function Linear
Regularization Parameter (C)	Controls the trade-off between achieving a low training error and a low testing error (controls overfitting and underfitting).	[0.1, 1, 10, 100]
Epsilon	A tube-like structure, established around datapoints, within which no penalty is given to errors in the SVR model. It defines the margin of tolerance for errors.	[0.1, 0.2, 0.5]
Gamma	It is a kernel function which influences the shape of the decision boundary and how much influence individual data points have.	'Scale' and 'auto' and range of values generated using np.logspace (-3, 3, 7)

c) Random Forest (RF)

Random Forest is an ensemble machine learning technique which combines multiple decision trees for prediction. The key hyperparameters of Random Forest, along with their respective initializations, are as follows:

Table 2 Parameter Description Table RF

Hyperparameter	Description	Initialization
n estimators	Represents the number of decision trees in the Random Forest.	[100, 200, 300]
Max depth	Controls the maximum depth of each decision tree.	[None, 10, 20 ,30]
Min samples split	Specifies the minimum number of samples required to split and internal node of a tree.	[2, 5, 10]
Min samples leaf	Specifies the minimum number of samples required to be at a leaf node.	[1, 2, 4]

3 Results and Discussion

3.1 Analysis of Discharge and Sediment Data

Figure 4 illustrates the temporal variation of discharge (blue line) and suspended sediment concentration (red dots) at Station 430.5 in the Seti-Gandaki River from 2004 to 2017, revealing distinct seasonal patterns characteristic of monsoon-influenced Himalayan River systems. The discharge data demonstrates pronounced annual cycles with peak flows exceeding 1,400 m³/s during monsoon periods (typically June-October) and low flows dropping to near-zero during dry seasons (November-May), while suspended sediment concentrations exhibit strong correlation with discharge patterns, reaching maximum values above 20,000 mg/L during high-flow events. The plot clearly shows the temporal data availability disparity mentioned in the study, with continuous discharge measurements throughout the entire period contrasting sharply with the sparse and irregular distribution of sediment concentration observations, particularly evident in the gaps during several years where sediment data are completely absent. This visualization effectively demonstrates the challenge of developing accurate sediment rating curves from incomplete datasets and underscores the necessity for machine learning approaches that can leverage the complete discharge record to predict sediment concentrations during data-scarce periods, while also highlighting the strong seasonal variability that necessitates season-specific modeling approaches for accurate sediment transport estimation.

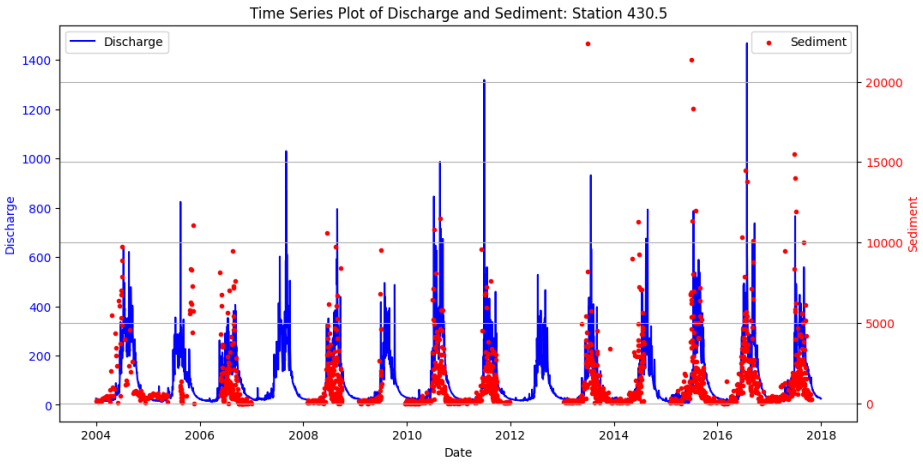

Figure 4: Time series plot of Discharge and Sediment

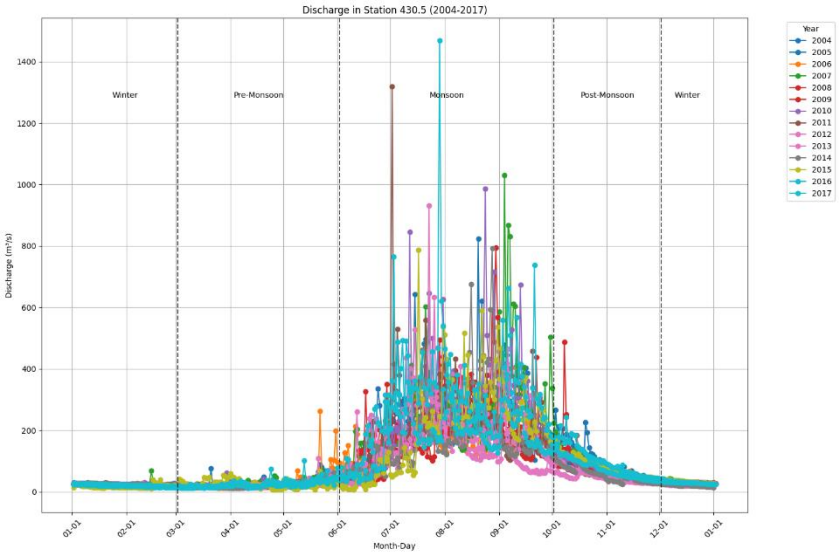

Figure 5: Discharge in Station 430.5

Figure 5 displays the seasonal discharge patterns at Station 430.5 from 2004-2017, clearly delineating the four distinct hydrological seasons in Nepal's river systems. The data demonstrates pronounced seasonal variability, with minimal flows (typically <50 m³/s) during winter months (December-March), moderate increases during the pre-monsoon period (April-June) due to snowmelt, dramatic peak discharges exceeding 1,400 m³/s during the monsoon season (June-October), and gradual recession during the post-monsoon period (October-December). The overlapping-colored lines representing different years reveal consistent seasonal patterns across the study period, though with notable inter-annual variability in peak discharge magnitudes, particularly evident during monsoon months where some years show significantly higher maximum flows than others. This visualization effectively illustrates the strong seasonality that characterizes Himalayan River systems and provides the foundation for understanding why season-specific sediment rating curve development is essential for accurate sediment transport modeling in such highly variable hydrological regimes.

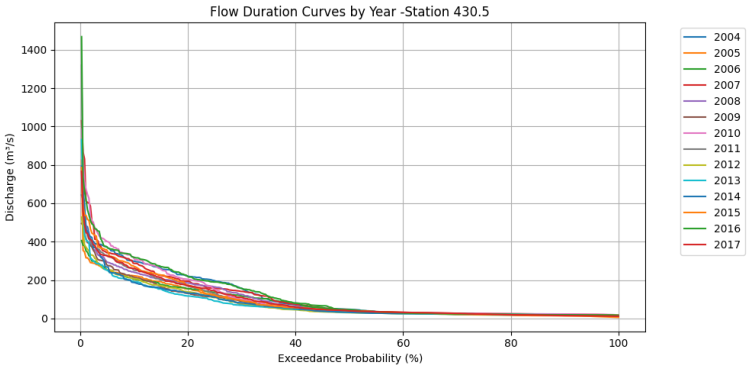

Figure 6: flow duration curves for Station 430.5

The flow duration curves for Station 430.5 as illustrated in Figure 6 overlay annual discharge ranked by exceedance probability for each year from 2004 to 2017, illustrating how often specific flow magnitudes are equaled or exceeded. All curves display the characteristic steep initial decline, indicating that high flows (e.g., >400 m³/s) occur infrequently (low exceedance probabilities), followed by a gentler slope through moderate flows and a long tail of low flows that persist for the majority of the year. Inter-annual variability is most pronounced at the high-flow end of the curves, where peak discharges in certain years (notably 2006 and 2013) substantially exceed those of other years, while low-flow characteristics remain relatively consistent across the study period. This plot underscores both the strong seasonal influence on flow regimes and the variability of extreme events, information that is critical for designing sediment rating curves and managing water resources under fluctuating hydrological conditions.

Figure 7 depicts daily suspended sediment concentrations at Station 430.5 over the 2004–2017 period, with data points colored by year and partitioned into Nepal's four hydrological seasons. Sediment levels remain near zero during winter (December–March) and rise moderately in the pre-monsoon (April–June) as snowmelt contributes to increased transport. The monsoon season (June–October) exhibits the highest and most frequent peak concentrations—exceeding 20 000 ppm in some years—reflecting intense rainfall-driven erosion and runoff, while the post-monsoon period (October–December) shows a rapid decline back to low levels. Inter-annual variability is evident in the magnitude and timing of peaks, illustrating how seasonal factors and extreme events influence sediment mobilization; this underscores the need for seasonally tailored sediment rating curves to accurately capture these dynamic patterns.

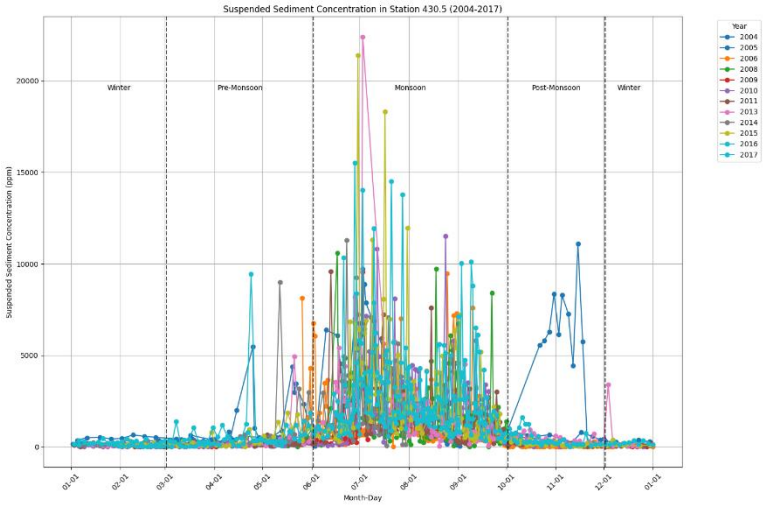

Figure 7: Sediment on station 430.5

This log scale scatterplot (Figure 8) illustrates the relationship between daily discharge and suspended sediment concentration at Station 430.5 from 2004–2017, with data points colored by month to highlight seasonal effects. Low discharges (<20 m³/s) during winter and early spring (purple–blue hues) correspond to minimal sediment concentrations (tens to hundreds of ppm), whereas higher discharges (50–500 m³/s) in the monsoon season (green–yellow hues) align with elevated sediment concentrations (hundreds to several thousand ppm). The upward-trending cloud of points demonstrates a clear positive correlation between discharge and sediment load, with greater scatter at peak flows reflecting variability in sediment availability and hysteresis effects during rapid hydrograph rises and falls. This pattern underscores the need for models that can capture the seasonally varying, non-linear discharge–sediment relationship for accurate sediment transport estimation.

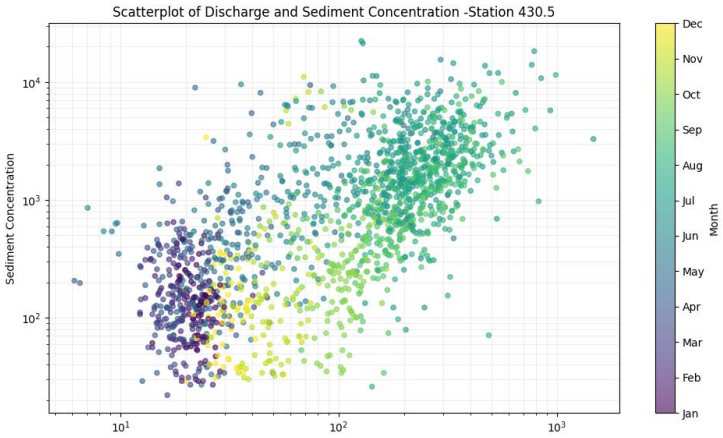

Figure 8: Scatterplot of Discharge and Sediment Concentration

3.2 Seasonal Variation

a) Pre-monsoon Analysis

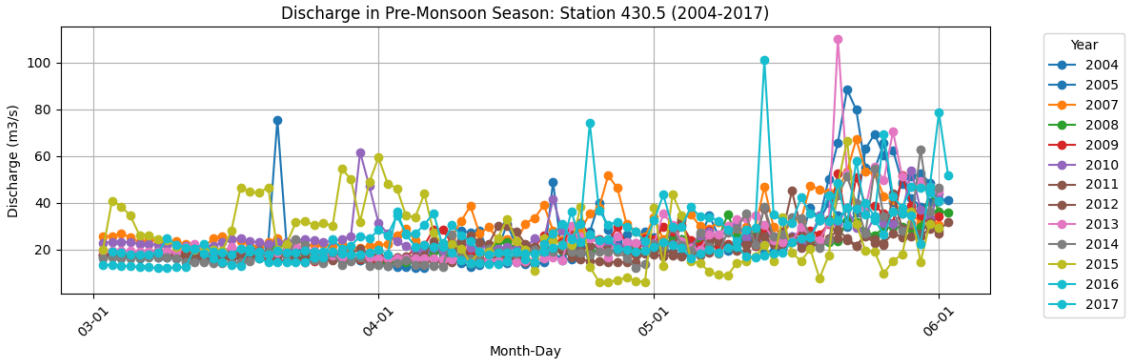

Figure 9: Discharge in Pre-monsoon season

Figure 9 presents daily discharge at Station 430.5 during the pre-monsoon season (March–May) from 2004 to 2017, showing consistently low flows (~10–30 m³/s) in March that gradually increase to ~40–60 m³/s by late May, driven by snowmelt and early rains. Superimposed on this seasonal rise are occasional sharp spikes exceeding 100 m³/s in certain years—most notably 2008 and 2017—indicating isolated hydrological events. This trend underscores the predictable seasonal transition in flow as well as the impact of episodic snowmelt or pre-monsoon storms on river discharge, information crucial for water resource planning in this period.

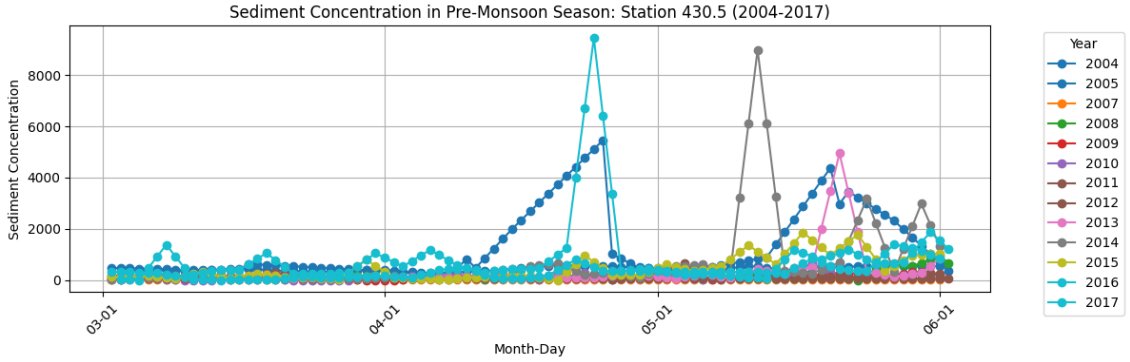

Figure 10: Sediment in Pre-monsoon season

The pre-monsoon suspended sediment plot as illustrated in Figure 10 for Station 430.5 (March–May, 2004–2017) shows low concentrations (~0–1,000 ppm) through March and early April, followed by a gradual rise in late April and May, with occasional sharp spikes up to ~9,000 ppm peaking in late May likely due to snowmelt or pre-monsoon rains. While most years maintain stable, low sediment levels, pronounced peaks in years like 2008 and 2017 reflect isolated hydrological events that sharply increase sediment transport. This pattern underscores the generally minimal sediment flux during pre-monsoon and the importance of episodic events in driving sediment dynamics, informing erosion control and sediment management strategies.

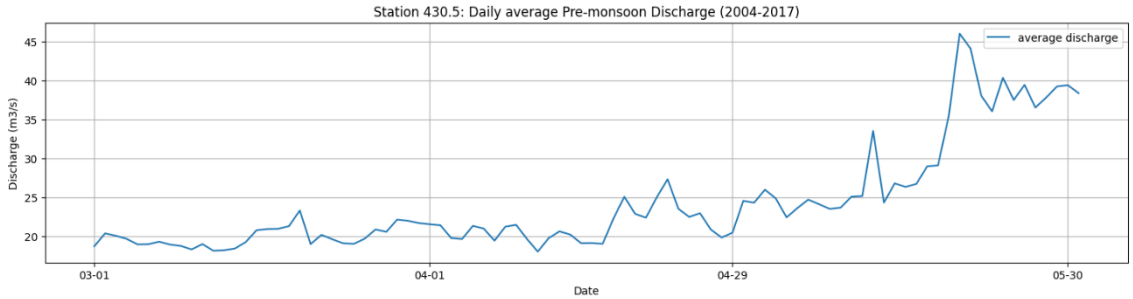

Figure 11: Daily Average Pre-monsoon Discharge

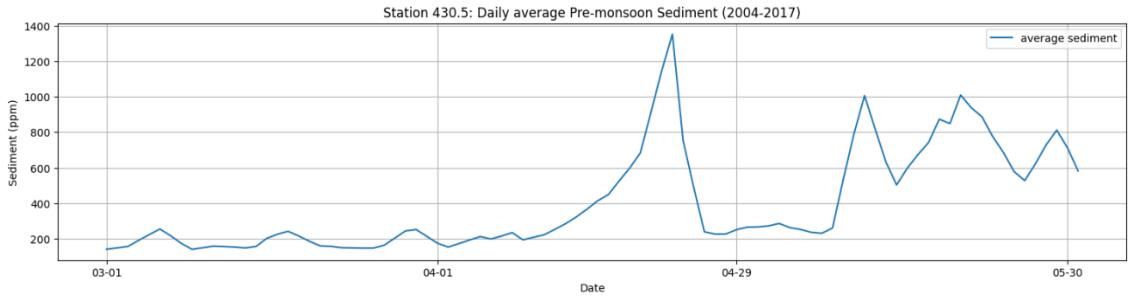

Figure 12: Daily Average Pre-monsoon Sediment

The paired plots of daily average pre-monsoon suspended sediment concentration and discharge (Figure 11 and 12) at Station 430.5 (2004–2017) show a clear seasonal rise: discharge remains relatively steady around 18–25 m³/s in March and April before climbing to 30–45 m³/s by late May, while sediment concentrations mirror this trend, holding near 200–300 ppm early on and then surging to peaks of 1 300 ppm at the end of April and around 1 000 ppm in late May. These synchronized increases indicate that enhanced snowmelt and early monsoon rains progressively boost both water flow and sediment transport, with the sharpest sediment spike coinciding with the steepest discharge rise, underscoring the tight coupling between hydrological forcing and sediment dynamics during the pre-monsoon transition.

The strong correlation (0.71) suggests that as discharge increases, sediment concentration also tends to increase significantly during the Pre-Monsoon season. This relationship reflects the influence of rising river

flow, driven by snowmelt and pre-monsoon rainfall, on sediment transport. This insight is critical for understanding sediment dynamics and managing river systems effectively during this transitional period.

b) Monsoon Analysis

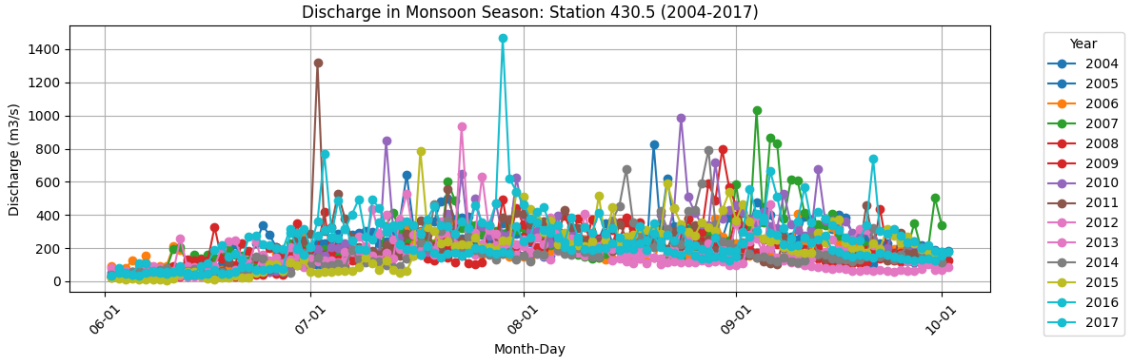

Figure 13: Discharge in Monsoon season

Figure 13 depicts daily discharge at Station 430.5 during the monsoon season (June–October) from 2004 to 2017, demonstrating the dramatic hydrological variability characteristic of Nepal's monsoon climate. Discharge begins around 50–100 m³/s in early June, then exhibits highly erratic patterns with frequent peaks exceeding 1,000 m³/s and occasional extreme spikes reaching 1,400+ m³/s during intense rainfall events. The highest discharges typically occur in July and August, with notable inter-annual variability years like 2006, 2011, and 2017 showing particularly pronounced flood peaks compared to drier years. This monsoon discharge pattern reflects the basin's response to intense, episodic rainfall events that drive the majority of annual water yield and create challenging conditions for flood management and infrastructure planning.

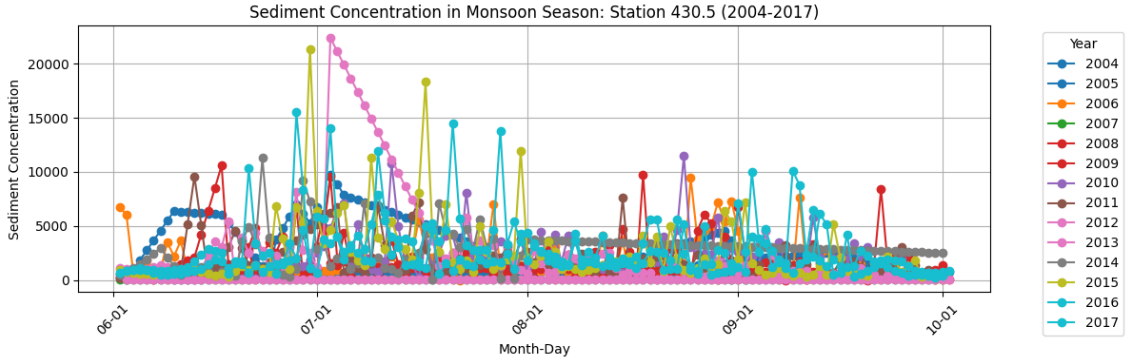

Figure 14: Sediment in Monsoon Season

The monsoon suspended sediment concentration plot (Figure 14) for Station 430.5 (June–October, 2004–2017) reveals extreme variability with concentrations ranging from near-zero during low-flow periods to exceptional peaks exceeding 20,000 ppm during major flood events. The highest sediment concentrations occur primarily in June and July, coinciding with early monsoon intensity, with notable spikes in years such as 2013 and 2015 that reflect particularly erosive rainfall-runoff events. Unlike discharge patterns that show sustained elevated flows, sediment concentrations exhibit sharp, transient pulses that rapidly decline even as flows remain high, indicating complex sediment supply limitations and hysteresis effects during prolonged monsoon periods. This pattern underscores the critical importance of early monsoon events in driving annual sediment transport and highlights the challenges of predicting sediment loads during this highly dynamic period.

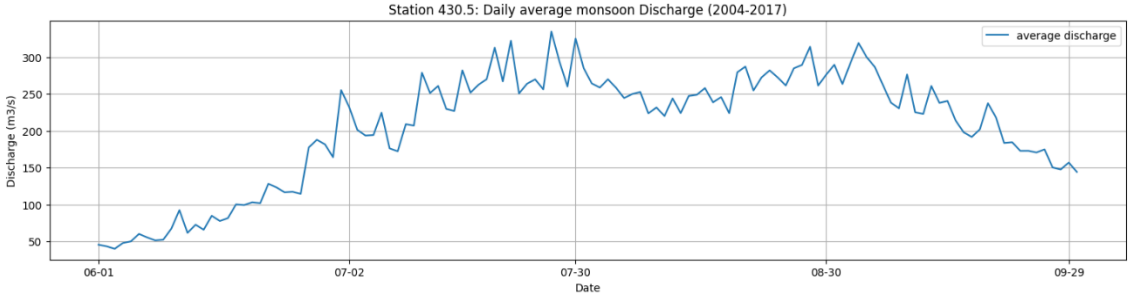

Figure 15: Daily Average Monsoon Discharge

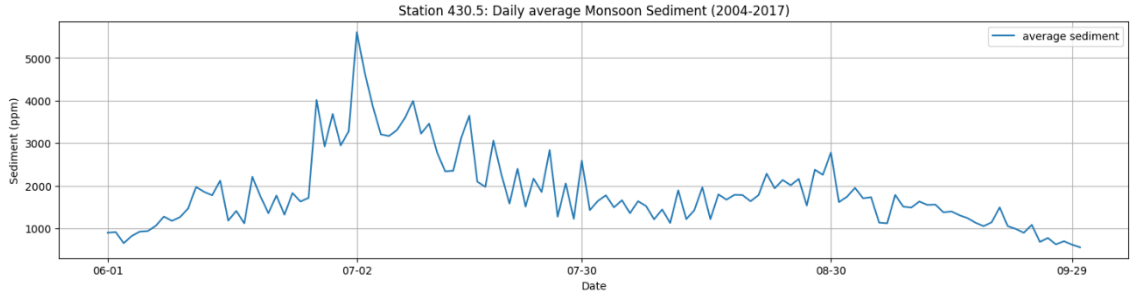

Figure 16: Daily Average Monsoon Sediment

Figures 15 and 16 present daily average monsoon discharge and suspended sediment patterns at Station 430.5 (2004–2017), revealing the seasonal evolution of the monsoon period. Average discharge shows a steady rise from ~50 m³/s in early June to peak values of ~300 m³/s by mid-July, followed by sustained high flows through August before gradually declining to ~150 m³/s by late September. Suspended sediment concentrations follow a more complex pattern, with the highest average values (~5,000 ppm) occurring in early July, followed by a gradual decline despite continued high discharge, suggesting sediment exhaustion effects and the dominance of early monsoon erosion in annual sediment budgets. The sediment peak preceding the discharge maximum illustrates the "first flush" phenomenon where initial monsoon rains mobilize accumulated sediments more effectively than sustained high flows, providing crucial insights for sediment management strategies and reservoir siltation predictions during the monsoon season.

The weak correlation (0.27) suggests that while there is some relationship between discharge and sediment concentration during the monsoon season, other factors such as localized erosion, rainfall intensity, and sediment availability may play a more significant role in determining sediment transport. This insight is critical for understanding sediment dynamics and planning river management strategies during the monsoon period.

c) Post-monsoon Analysis

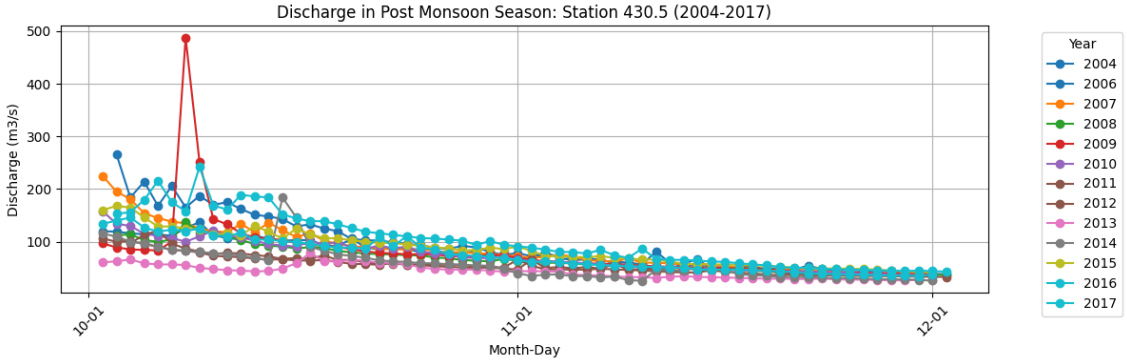

Figure 17: Discharge in Post-Monsoon season

Figure 17 illustrates daily discharge at Station 430.5 during the post-monsoon season (October–December) from 2004 to 2017, showing a characteristic recession pattern as the basin transitions from monsoon peak flows to winter base flows. Discharge begins at relatively elevated levels (~100–200 m³/s) in early October, reflecting the tail end of monsoon runoff, before exhibiting a steady exponential decline to ~50–80 m³/s by late December. Notable exceptions include occasional spikes, particularly the extreme peak approaching 500 m³/s in 2009, likely attributed to late-season storm events or delayed monsoon effects. The overall trend demonstrates the basin's gradual return to base flow conditions, with most years following a similar recession curve despite some inter-annual variability in the rate of decline, providing predictable conditions crucial for post-monsoon water resource planning and infrastructure maintenance.

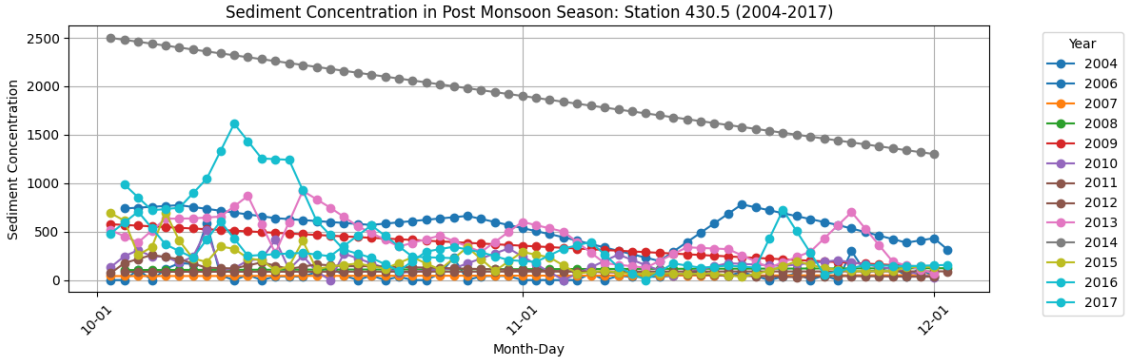

Figure 18: Sediment in Post-Monsoon season

The post-monsoon suspended sediment concentration plot (Figure 18) for Station 430.5 (October–December, 2004–2017) reveals a sharp decline from residual monsoon sediment levels to minimal winter concentrations. Early October shows elevated concentrations (~500–2,500 ppm) as delayed monsoon effects continue to mobilize sediments, with 2014 displaying particularly high initial values approaching 2,500 ppm. However, sediment concentrations rapidly decrease through October and November, stabilizing at low levels (~100–300 ppm) by December as sediment sources become exhausted and flows diminish. Unlike the monsoon period's extreme variability, post-monsoon sediment transport exhibits relatively predictable recession patterns with occasional spikes likely corresponding to late-season rainfall events. This pattern reflects the basin's transition from active erosion to minimal sediment mobilization, with the majority of annual sediment transport completed during the preceding monsoon period.

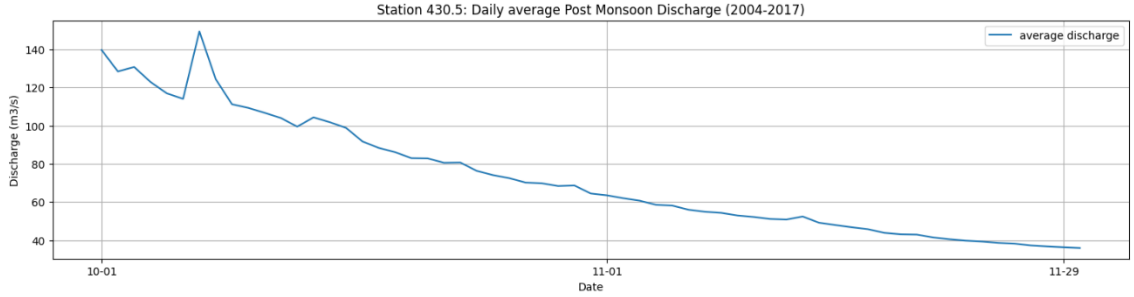

Figure 19: Daily Average Discharge in Post-Monsoon Season

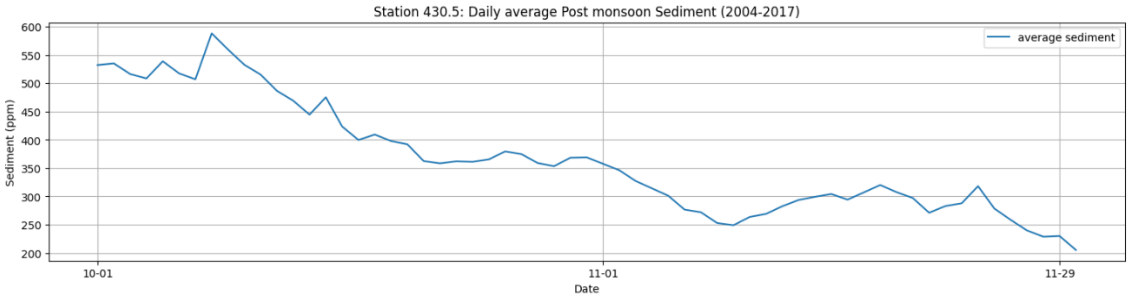

Figure 20: Daily Average Sediment in Post-Monsoon Season

Figures 19 and 20 present daily average post-monsoon patterns at Station 430.5 (2004–2017), demonstrating the synchronized recession of both discharge and sediment concentration during this transitional period. Average discharge shows a smooth exponential decline from ~45 m³/s in early October to ~20 m³/s by late December, while suspended sediment concentrations follow a similar but steeper recession from peak values of ~580 ppm in early October to ~200 ppm by late November. The sediment recession curve exhibits a more pronounced early decline compared to discharge, suggesting rapid depletion of readily available sediment sources following monsoon mobilization. This pattern indicates that post-monsoon sediment transport is primarily driven by the washout of residual monsoon-mobilized sediments rather than active erosion processes, making this period relatively predictable for sediment management and providing stable conditions for infrastructure maintenance and water treatment operations.

The strong correlation (0.95) between post-monsoon discharge and sediment concentration reflects the relatively stable relationship during this recession period, where both variables follow predictable exponential decay patterns. This relationship indicates that traditional sediment rating curves may be more applicable during post-monsoon conditions compared to the complex dynamics of monsoon and pre-monsoon periods,

though the rapid decline in both variables limits the practical duration of high sediment transport during this season.

d) Winter Analysis

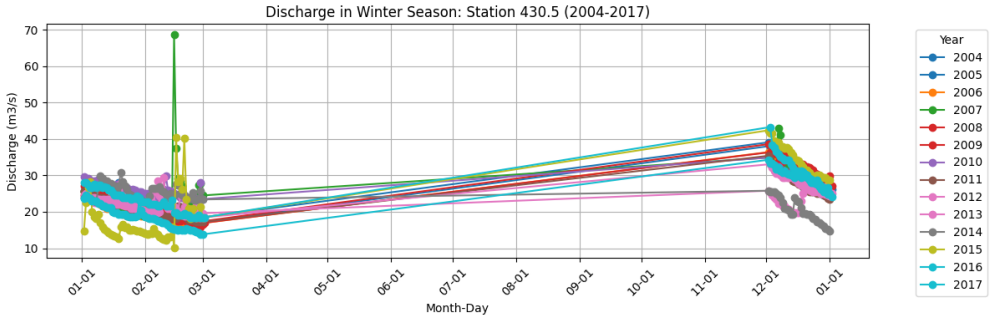

Figure 21: Discharge in Winter season

Figure 21 displays daily discharge at Station 430.5 during the winter season (December–March) from 2004 to 2017, illustrating the low-flow regime characteristic of Nepal's dry season. Discharge remains consistently low throughout most of the winter period, ranging between 15–30 m³/s with minimal day-to-day variability, reflecting the dominance of groundwater baseflow and limited surface runoff due to reduced precipitation and frozen conditions at higher elevations. Notable exceptions include isolated spikes reaching up to 70 m³/s, particularly evident in early winter periods of certain years like 2006, likely attributed to delayed recession from post-monsoon flows or occasional winter storm events. The gradual increase observed toward late February and March signals the onset of pre-monsoon snowmelt contributions, providing a smooth transition from winter base flows to the more dynamic pre-monsoon period and demonstrating the predictable nature of winter discharge patterns essential for dry-season water resource allocation.

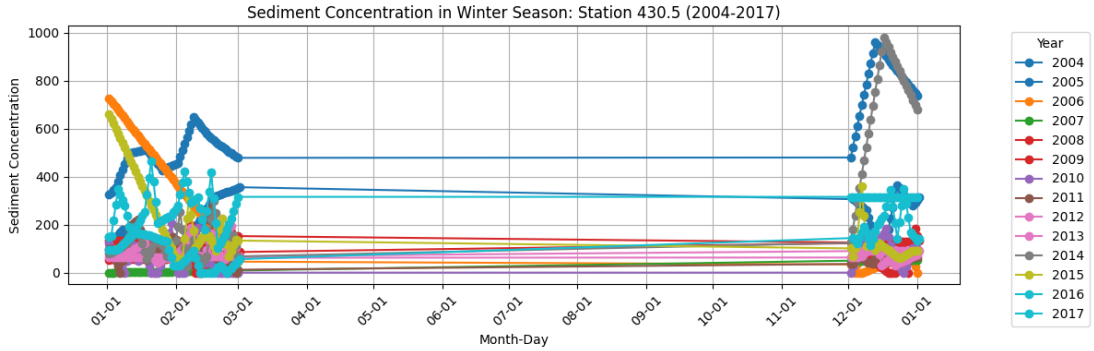

Figure 22: Sediment in Winter season

The winter suspended sediment concentration plot (Figure 22) for Station 430.5 (December–March, 2004–2017) reveals the basin's minimal sediment transport capacity during the dry season, with concentrations typically remaining below 200 ppm throughout most of the winter period. Early winter shows slightly elevated concentrations (~200–400 ppm) representing the final washout of post-monsoon sediments, before settling into consistently low levels that reflect limited erosion potential under low-flow conditions. Exceptional spikes approaching 1,000 ppm in certain years (notably 2004 and 2005) likely correspond to isolated winter precipitation events or localized disturbances, while the dramatic increase observed in late February/early March signals the beginning of snowmelt-driven sediment mobilization. This pattern demonstrates the seasonal minimum in sediment transport, with winter conditions providing the most stable and predictable sediment regime of the annual cycle, crucial for water treatment planning and infrastructure maintenance during the dry season.

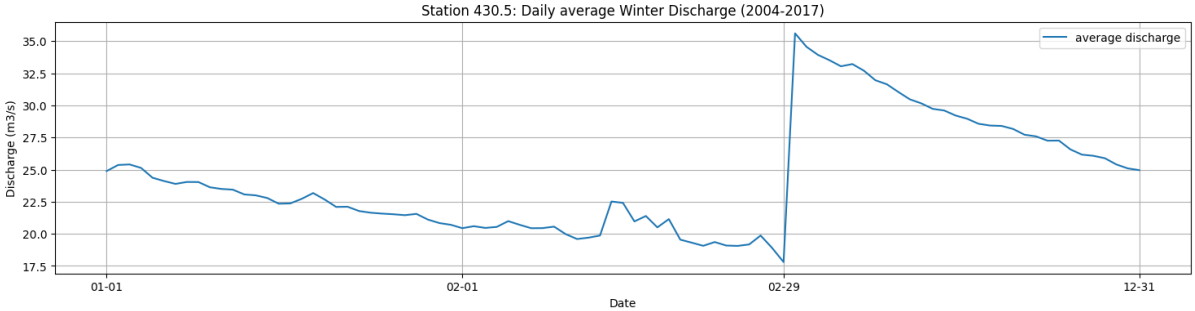

Figure 23: Daily Average Discharge in Winter Season

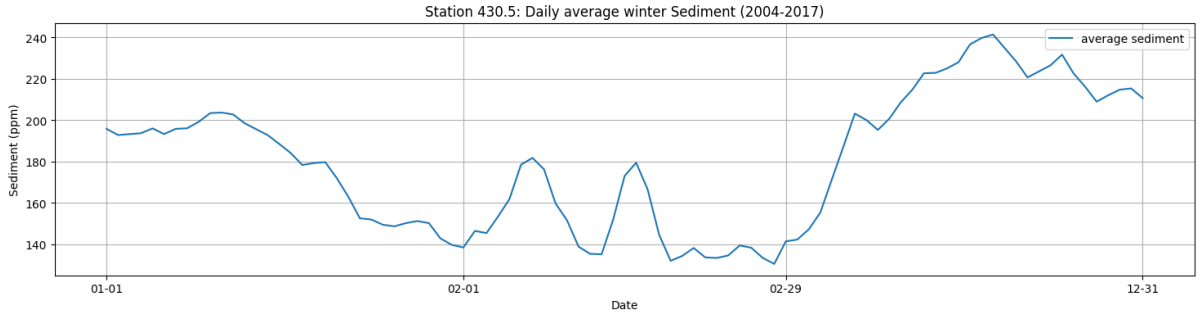

Figure 24: Daily Average Sediment in Winter Season

Figures 23 and 24 present daily average winter discharge and suspended sediment patterns at Station 430.5 (2004–2017), revealing the stable characteristics of the dry season followed by the transition to pre-monsoon conditions. Average discharge shows remarkably consistent values around 20–25 m³/s from December through mid-February, with a pronounced increase beginning in late February and continuing through March, reaching ~35 m³/s by month's end as snowmelt contributions intensify. Suspended sediment concentrations exhibit a similar but more variable pattern, starting near 200 ppm in early winter, dropping to seasonal lows (~140 ppm) in mid-February, then sharply increasing to ~240 ppm by late March as warming temperatures enhance sediment mobilization. The concurrent rise in both discharge and sediment toward winter's end illustrates the synchronized response to seasonal warming and marks the hydrological transition toward the more dynamic pre-monsoon period, providing critical insights for seasonal water resource management and the timing of infrastructure maintenance activities.

The moderate correlation (0.61) between winter discharge and sediment concentration reflects the generally stable relationship during low-flow conditions, where limited variability in both parameters results in predictable sediment transport patterns. This relationship suggests that traditional sediment rating approaches may be most applicable during winter months, though the overall low magnitude of sediment transport limits the practical significance of this period for annual sediment budgets while providing optimal conditions for detailed sediment process studies and model calibration.

3.3 Processed Daily Average Data

To enhance data quality and strengthen the reliability of seasonal sediment-discharge relationships, a systematic outlier detection and removal procedure was implemented on the daily average sediment and discharge datasets across all four hydrological seasons: pre-monsoon, monsoon, post-monsoon, and winter. The Interquartile Range (IQR) method was employed to identify anomalous values, whereby data points falling below the first quartile minus 1.5 times the IQR ($Q1 - 1.5 \times IQR$) or exceeding the third quartile plus 1.5 times the IQR ($Q3 + 1.5 \times IQR$) were classified as outliers and subsequently excluded from further analysis. Figures 25 to 32 present the filtered daily average discharge and sediment datasets, demonstrating the improved data consistency achieved through this quality control process and providing a robust foundation for developing accurate seasonal sediment rating curves.

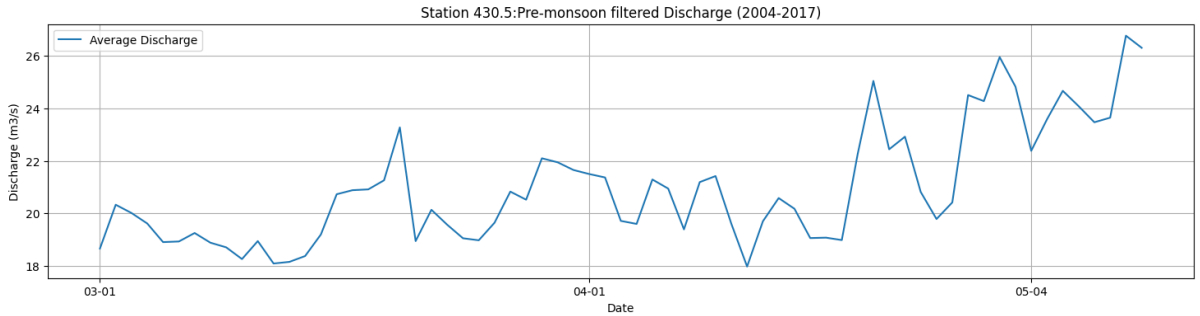

Figure 25: Pre-monsoon filtered Discharge

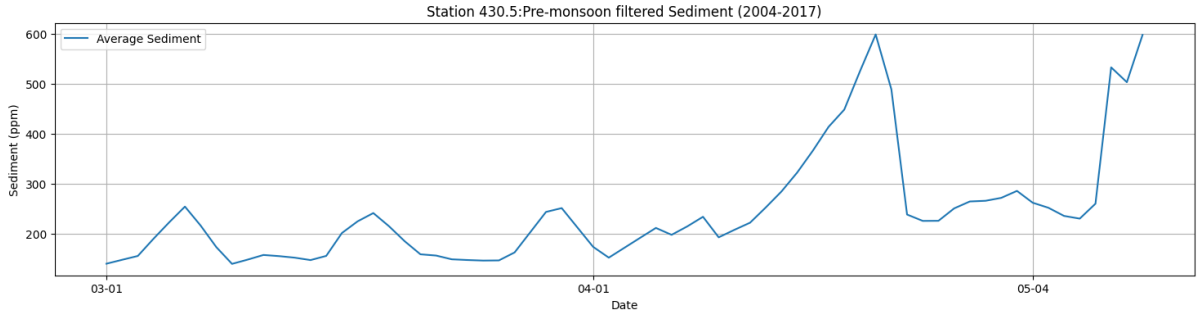

Figure 26: Pre-monsoon filtered Sediment

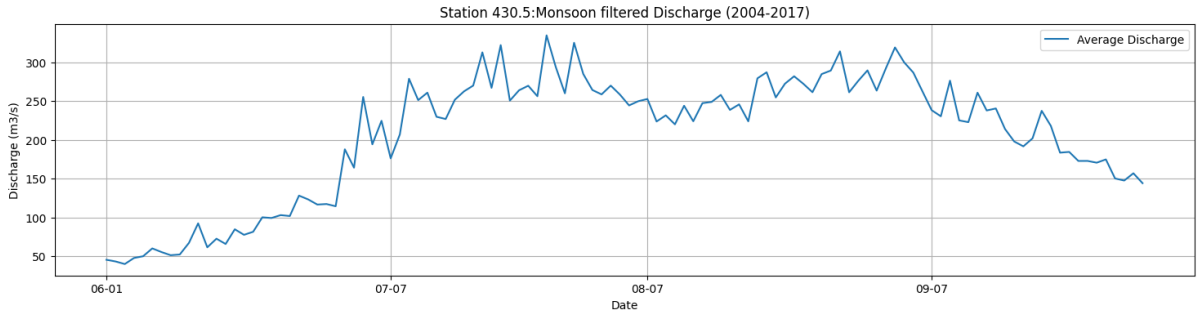

Figure 27: Monsoon filtered Discharge

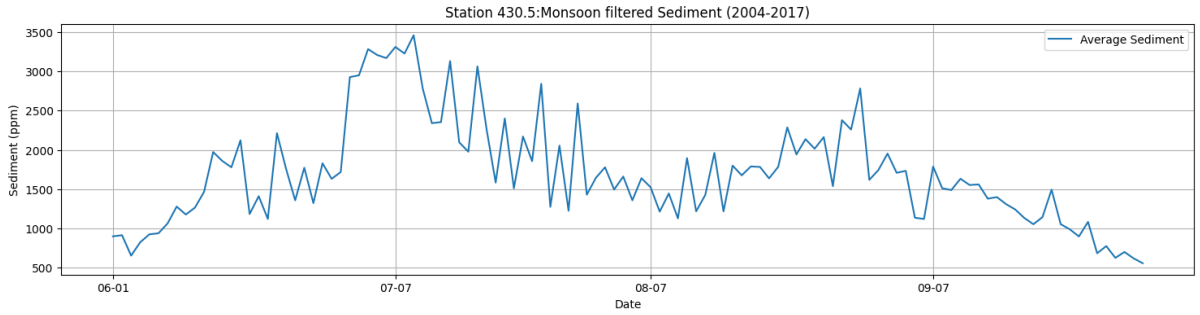

Figure 28: Monsoon filtered Sediment

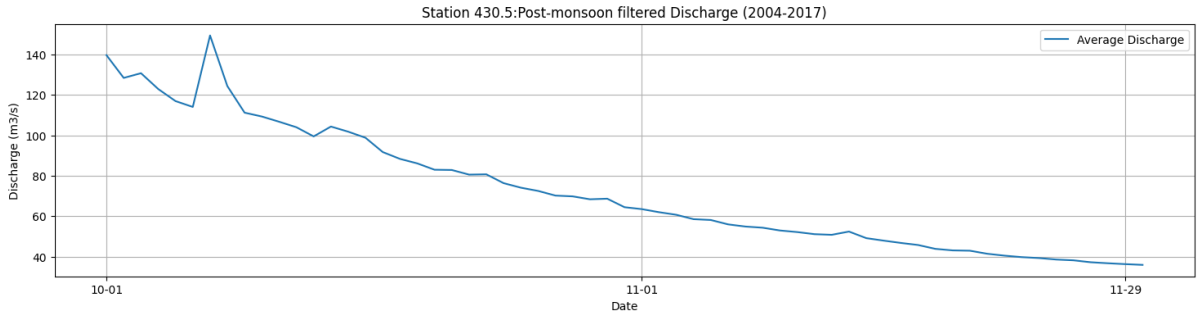

Figure 29: Post-monsoon filtered Discharge

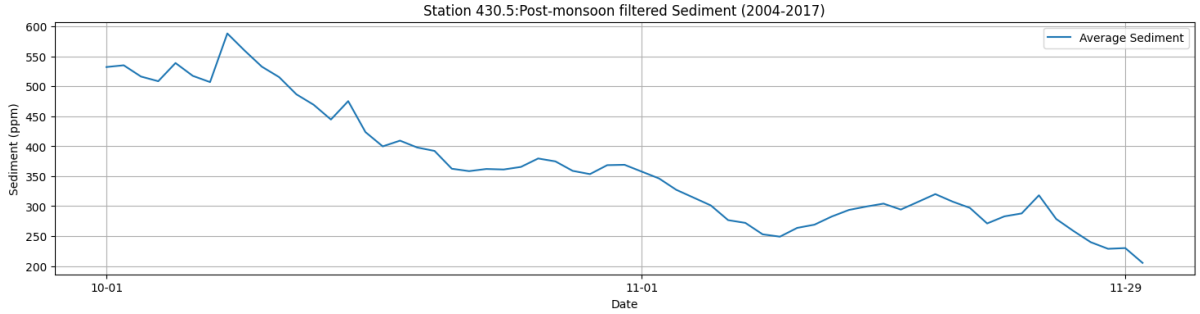

Figure 30: Post-monsoon filtered Sediment

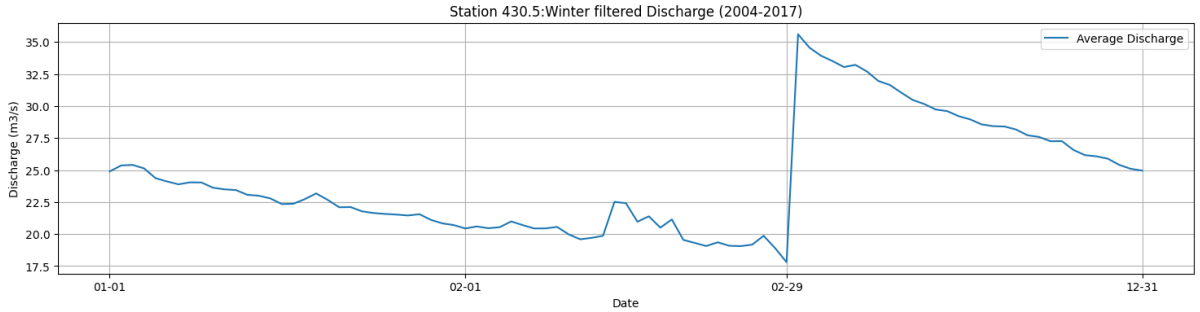

Figure 31: Winter filtered Discharge

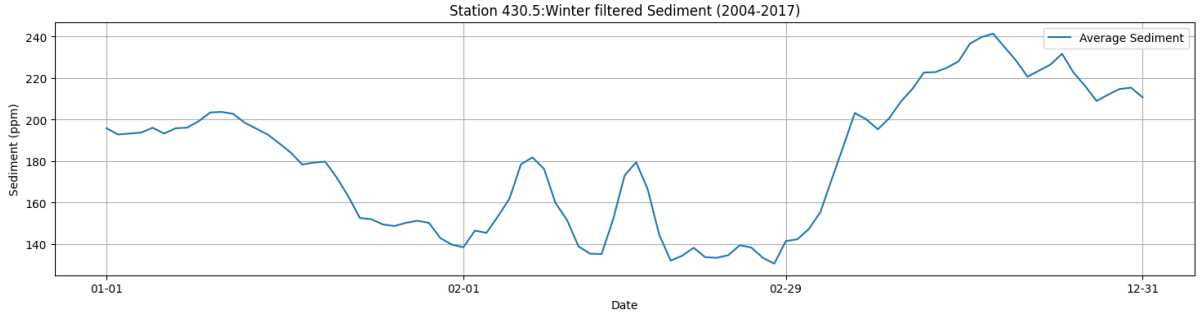

Figure 32: Winter filtered Sediment

3.4 Model Development

Three different machine learning models; KNN, SVM and Random Forest has been developed for each hydrological season with findings as follows:

a) KNN Pre-monsoon

The K-Nearest Neighbors (KNN) regression model performance for pre-monsoon sediment prediction at Station 430.5 demonstrates moderate predictive capability, as illustrated by the scatter plot showing actual versus predicted values (Figure 33) and the performance evaluation plot. The model achieves an R^2 value of

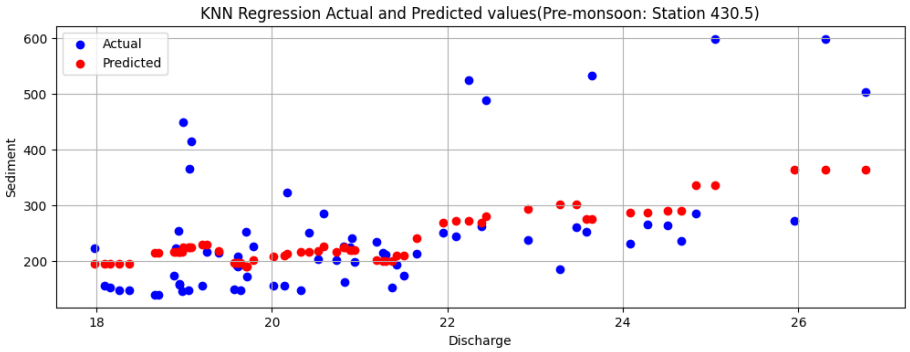

Figure 33: KNN model for pre-monsoon season

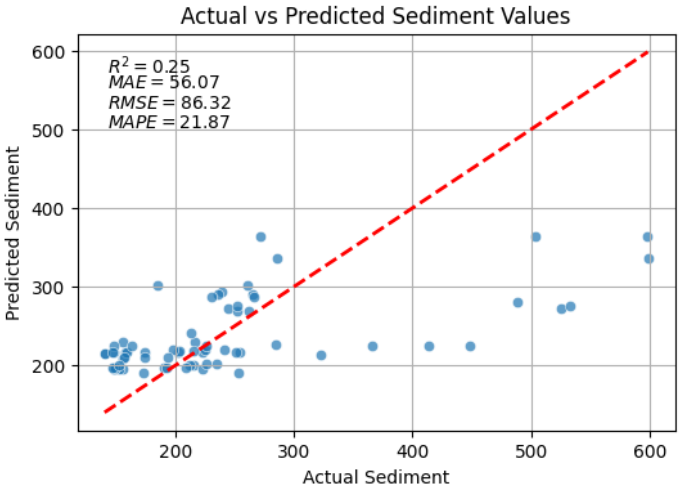

Figure 34: Scatter plot for KNN model -Pre-monsoon Season

0.25, indicating that approximately 25% of the variance in sediment concentration is explained by the discharge-based predictions, while the remaining 75% reflects the complex, non-linear relationships and additional factors influencing sediment transport during the pre-monsoon period. The performance metrics reveal a Mean Absolute Error (MAE) of 56.07 ppm, Root Mean Square Error (RMSE) of 86.32 ppm, and Mean Absolute Percentage Error (MAPE) of 21.87%, suggesting reasonable prediction accuracy for moderate sediment concentrations but with notable scatter, particularly evident in the actual vs. predicted plot where higher sediment values (>400 ppm) show considerable deviation from the 1:1 line. The regression plot (Figure 34) demonstrates that while the model captures the general trend of increasing sediment concentration with discharge, it struggles to predict extreme values and exhibits systematic underestimation of peak sediment concentrations, highlighting the challenges inherent in modeling pre-monsoon sediment transport dynamics using simple nearest-neighbor approaches.

b) KNN Monsoon

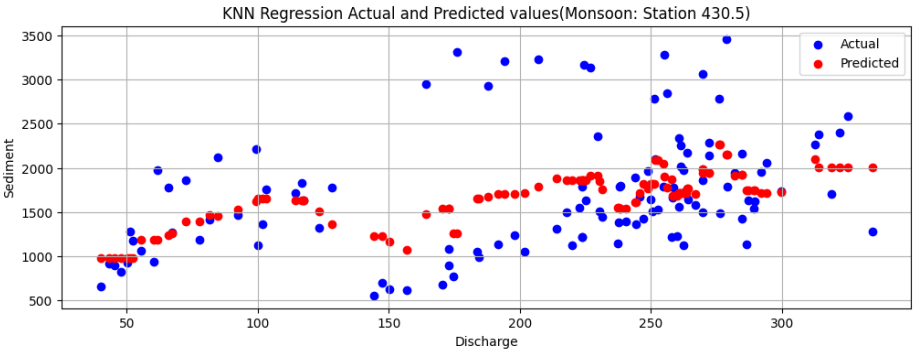

Figure 35: KNN model for Monsoon Season

The K-Nearest Neighbors (KNN) regression model for monsoon sediment prediction at Station 430.5 exhibits limited predictive performance, as demonstrated by the scatter plots showing substantial discrepancies between actual and predicted values across the full range of sediment concentrations. The model achieves a low R^2 value of 0.22, indicating that only 22% of the variance in sediment concentration is captured by the discharge-based predictions, while the performance metrics reveal considerable prediction errors with MAE of 418.87 ppm, RMSE of 543.99 ppm, and MAPE of 22.01%. The regression plot (Figure 36) illustrates the model's inability to accurately predict extreme sediment concentrations, particularly failing to capture peak values exceeding 3,000 ppm and showing systematic underestimation throughout the higher concentration ranges, while the actual vs. predicted plot demonstrates significant scatter around the 1:1 line with

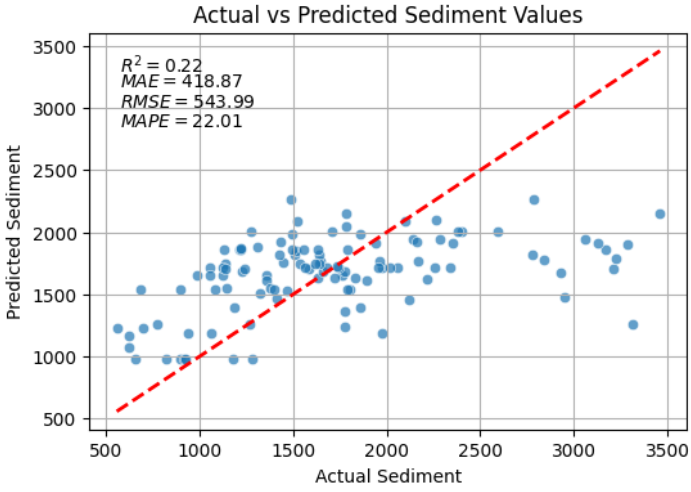

Figure 36: Scatterplot of KNN model -Monsoon Season

predictions generally constrained to a narrow range (1,000-2,500 ppm) despite actual values spanning from 500 to over 3,500 ppm. This poor performance reflects the complex, non-linear nature of monsoon sediment transport dynamics, where factors beyond simple discharge relationships including rainfall intensity, antecedent conditions, sediment availability, and hysteresis effects play crucial roles that the KNN algorithm struggles to capture effectively.

c) KNN Post-monsoon

The K-Nearest Neighbors (KNN) regression model demonstrates excellent predictive performance for post-monsoon sediment concentration at Station 430.5, as evidenced by the close alignment between actual and predicted values across the discharge range and the strong linear relationship in the performance evaluation plot. The model achieves an outstanding R^2 value of 0.97, indicating that 97% of the variance in sediment concentration is successfully captured by the discharge-based predictions, while the performance metrics show exceptional accuracy with low error values: MAE of 12.81 ppm, RMSE of 17.26 ppm, and MAPE of 3.35%. The regression plot reveals that predicted values (red dots) closely track actual values (blue dots) across the entire discharge spectrum from 40 to 140 m^3/s, with minimal scatter and consistent performance at both low and high sediment concentrations, while the actual vs. predicted plot shows data points clustering tightly around the 1:1 line with very little deviation. This superior performance reflects the more predictable and stable nature of post-monsoon sediment transport dynamics, where the exponential recession of both discharge and sediment concentration creates a strong, consistent relationship that the KNN algorithm can effectively capture, making this period particularly well-suited for traditional regression-based modeling approaches.

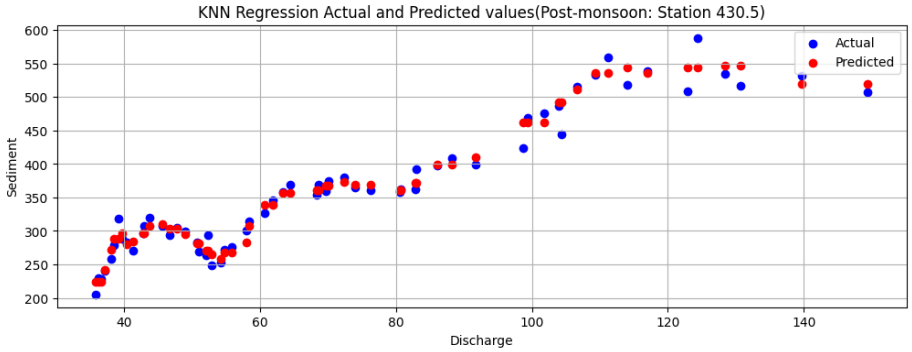

Figure 37: KNN model for Post-monsoon season

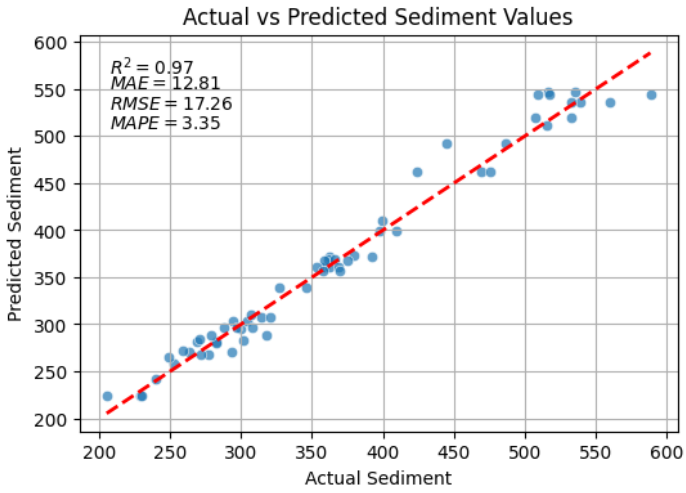

Figure 38: Scatterplot for KNN model -Post-monsoon Season

d) KNN winter

The K-Nearest Neighbors (KNN) regression model shows good predictive performance for winter sediment concentration at Station 430.5, demonstrating consistent accuracy

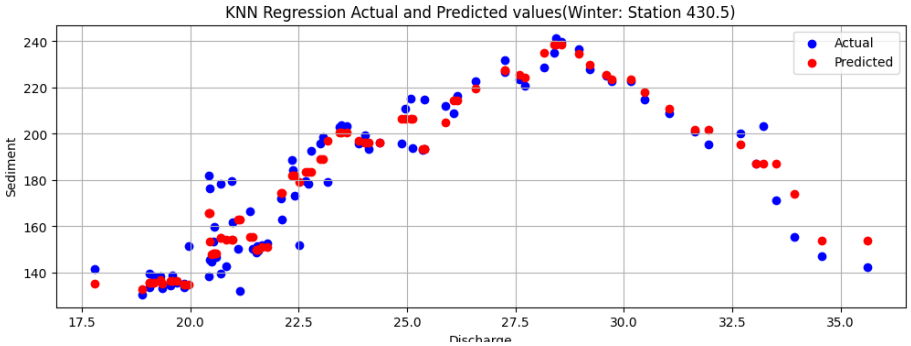

Figure 39: KNN model for Winter Season

across the relatively narrow range of discharge and sediment values characteristic of the dry season. The model achieves a solid R^2 value of 0.74, indicating that 74% of the variance in sediment concentration is explained by discharge-based predictions, with performance metrics showing reasonable accuracy: MAE of 11.01 ppm, RMSE of 14.07 ppm, and MAPE of 6.79%. The regression plot reveals that predicted values (red dots) closely follow actual values (blue dots) across the discharge range from 17.5 to 35 m^3/s, with particularly good agreement in the mid-range discharge values (20-30 m^3/s) where most winter data cluster, while the actual vs. predicted plot shows data points generally aligned along the 1:1 line with some scatter at the lower and higher ends of the sediment concentration range. This strong performance reflects the stable, low-flow conditions during winter that create more predictable discharge-sediment relationships compared to the complex dynamics of monsoon periods, making the KNN algorithm effective at capturing the relatively consistent patterns in winter sediment transport where baseflow conditions dominate and extreme variability is minimal

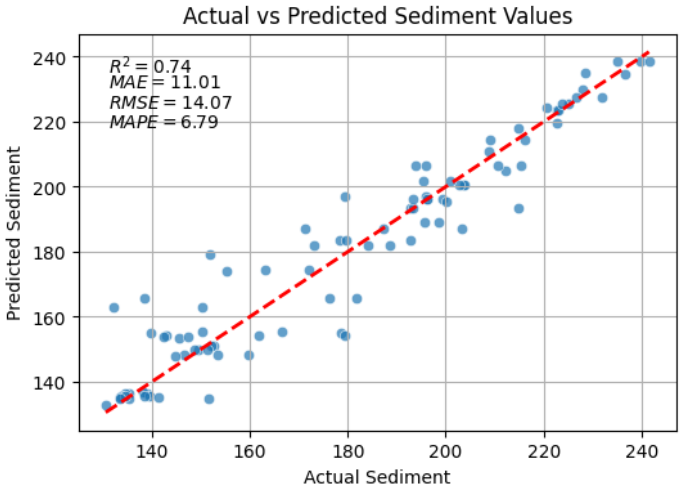

Figure 40: Scatter plot for KNN model -Winter Season

e) SVM Pre-monsoon

The Support Vector Machine (SVM) regression model for pre-monsoon sediment prediction at Station 430.5 demonstrates poor predictive performance, as evidenced by the significant scatter and systematic bias in both the regression and performance evaluation plots. The model achieves an extremely low R^2 value of 0.09, indicating that only 9% of the variance in sediment concentration is explained by the discharge-based predictions, representing a substantial deterioration compared to the KNN model's 25% variance explanation for the same season. The performance metrics reveal considerable prediction errors with MAE of 58.20 ppm, RMSE of 95.44 ppm, and MAPE of 19.43%, while the regression plot shows that predicted values (red dots) fail to capture the wide range of actual sediment concentrations, particularly struggling with higher values above 400 ppm where the model consistently underestimates sediment

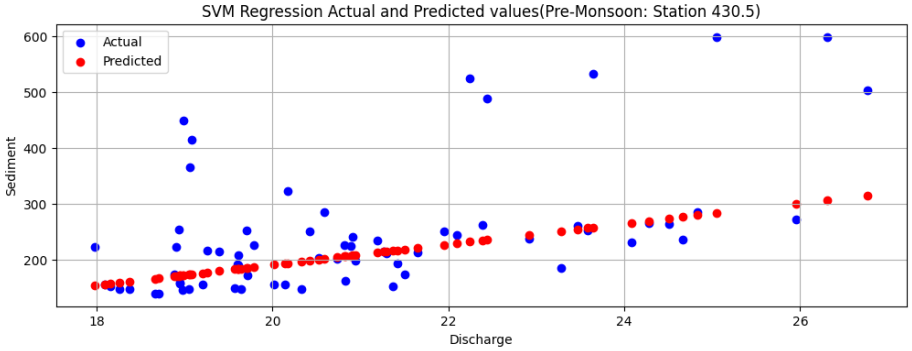

Figure 41: SVM model for Pre-monsoon season

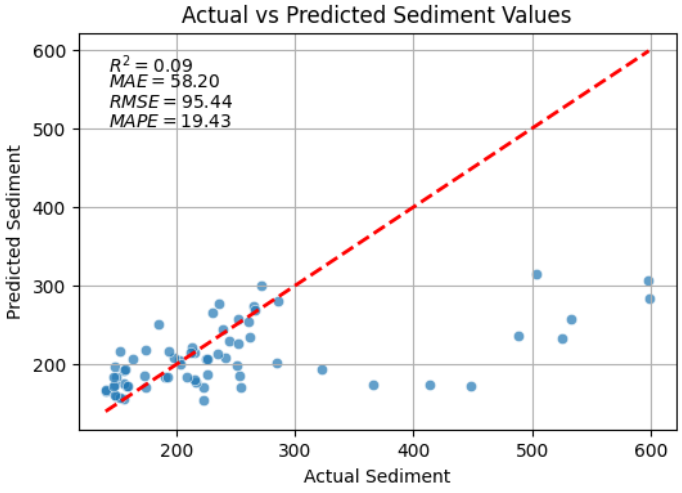

Figure 42: Scatterplot for SVM model -Pre-monsoon Season

loads. The actual vs. predicted plot demonstrates severe limitations in the SVM's ability to model pre-monsoon sediment dynamics, with predictions constrained to a narrow band between 150-320 ppm despite actual values ranging from approximately 150-600 ppm, and significant deviation from the 1:1 line throughout the concentration range. This poor performance suggests that the SVM model, despite its theoretical advantages in handling non-linear relationships through kernel functions, is unable to effectively capture the complex discharge-sediment interactions characteristic of the pre-monsoon period, possibly due to insufficient training data or inappropriate hyperparameter selection for this transitional hydrological season.

f) SVM Monsoon

The Support Vector Machine (SVM) regression model exhibits severely limited predictive capability for monsoon sediment concentration at Station 430.5, demonstrating one of the poorest performances across all seasonal models evaluated. The model achieves an unacceptably low R^2 value of 0.14, indicating that only 14% of the variance in sediment concentration is captured by discharge-based predictions, while the performance metrics reveal substantial prediction errors with MAE of 501.55 ppm, RMSE of 658.05 ppm, and MAPE of 27.09%. The regression plot illustrates the model's fundamental inability to respond to the extreme variability characteristic of monsoon sediment transport, with predicted values (red dots) forming a relatively flat, constrained band across the discharge spectrum, failing entirely to capture the dramatic peaks exceeding 3,000 ppm observed in actual data (blue dots). The actual vs. predicted plot reveals the extent of this predictive failure, showing that while actual sediment concentrations span from 500 to over 3,500 ppm, the SVM predictions are artificially constrained to a narrow range of approximately 1,000-1,900 ppm, creating systematic and severe underestimation of high sediment loads that dominate monsoon transport dynamics. This poor performance reflects the SVM model's inability to handle the extreme non-linearity, temporal variability, and complex hysteresis effects that characterize monsoon sediment transport, where multiple interacting factors beyond simple discharge relationships—including rainfall intensity, sediment supply limitations, and antecedent moisture conditions—govern sediment mobilization in ways that the current SVM configuration cannot adequately represent.

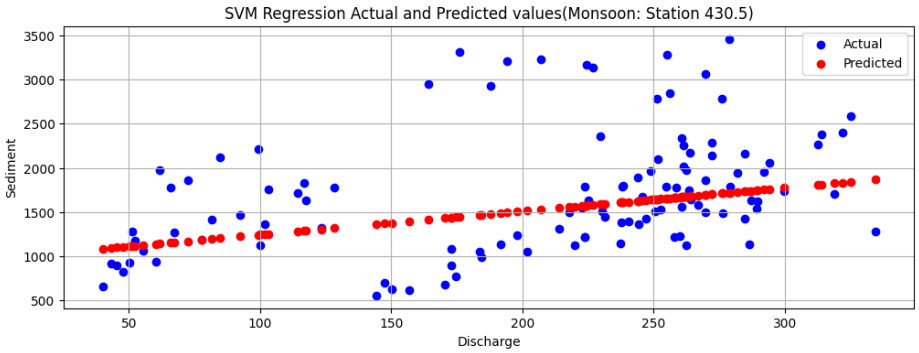

Figure 43: SVM model for Monsoon Season

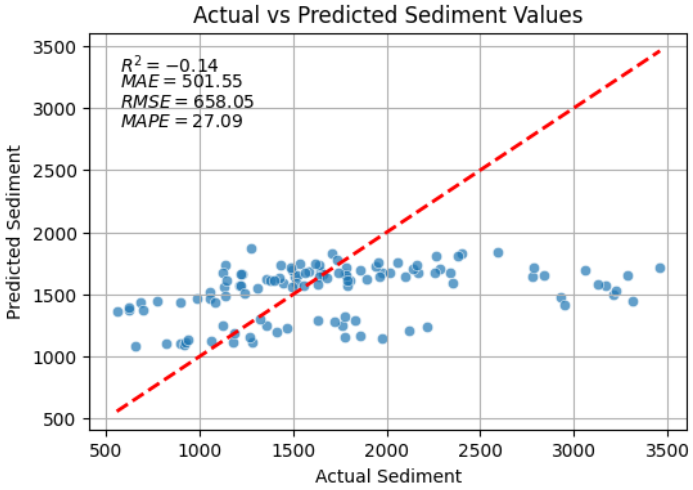

Figure 44: Scatterplot of SVM model -Monsoon Season

g) SVM Post-monsoon

The Support Vector Machine (SVM) regression model demonstrates excellent predictive performance for post-monsoon sediment concentration at Station 430.5, achieving results comparable to the outstanding KNN

performance for the same season. The model attains an impressive R^2 value of 0.93, indicating that 93% of the variance in sediment concentration is successfully captured by the discharge-based predictions, with performance metrics showing exceptional accuracy: MAE of 19.56 ppm, RMSE of 25.46 ppm, and MAPE of 4.99%. The regression plot reveals that predicted values (red dots) closely align with actual values (blue dots) across the entire discharge spectrum from 40 to 140 m^3/s, demonstrating consistent tracking throughout the post-monsoon recession period, while the actual vs. predicted plot shows data points clustering tightly around the 1:1 line with minimal scatter and strong linear correlation. This superior performance reflects the SVM model's ability to effectively capture the stable, exponential decay patterns characteristic of post-monsoon sediment transport, where both discharge and sediment concentration follow predictable recession curves that create strong, consistent relationships well-suited to the SVM's kernel-based approach. The model's success during this period demonstrates that SVM algorithms can excel when applied to hydrological regimes with clear, systematic patterns, making post-monsoon conditions particularly amenable to support vector machine modeling approaches.

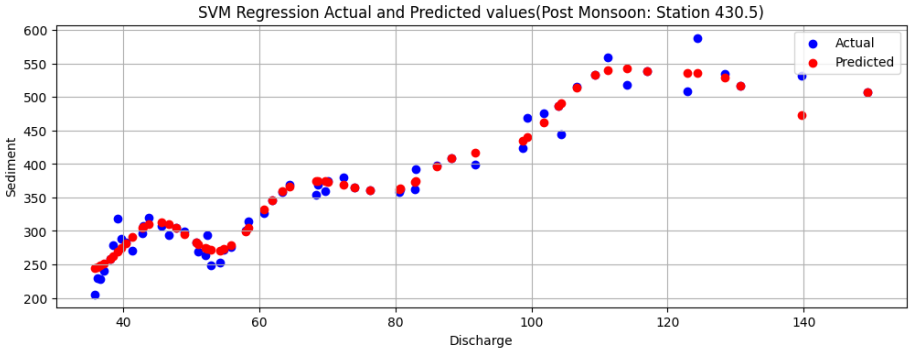

Figure 45: SVM model for Post-monsoon Season

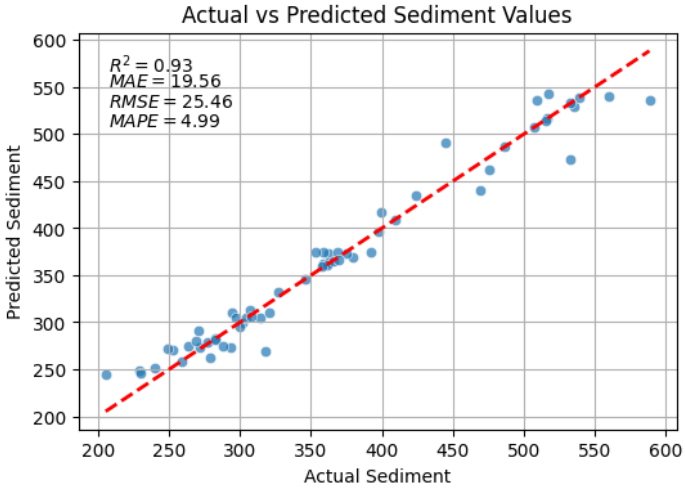

Figure 46: Scatterplot for SVM model -Post-monsoon Season

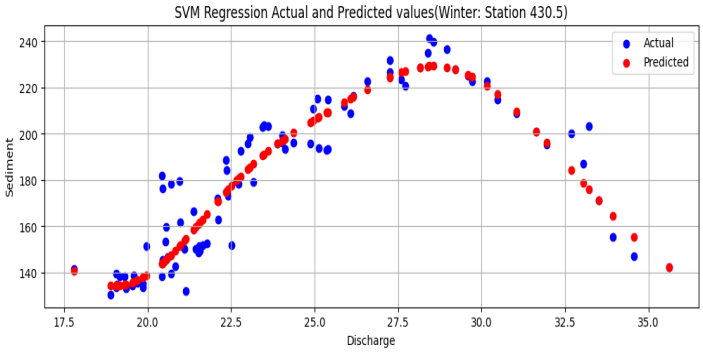

Figure 47: SVM model for Winter Season

h) SVM Winter

The Support Vector Machine (SVM) regression model shows strong predictive performance for winter sediment concentration at Station 430.5, demonstrating consistent accuracy across the stable, low-flow conditions characteristic of the dry season. The model achieves a solid R^2 value of 0.81, indicating that 81% of the variance in sediment concentration is explained by discharge-based predictions, with performance metrics showing good accuracy: MAE of 10.28

ppm, RMSE of 12.07 ppm, and MAPE of 6.28%. The regression plot reveals that predicted values (red dots) closely follow actual values (blue dots) across the discharge range from 17.5 to 35 m³/s, with particularly strong agreement in the central discharge range where most winter data cluster, while the actual vs. predicted plot shows data points well-aligned along the 1:1 line with relatively minor scatter at the extremes of the sediment concentration range. This strong performance reflects the SVM model's effectiveness in capturing the stable baseflow-dominated conditions during winter, where limited variability in both discharge and sediment concentration creates predictable relationships that the support vector approach can model successfully. The winter season's consistent patterns and minimal extreme events provide optimal conditions for SVM regression, demonstrating that the algorithm performs well when applied to hydrological regimes with reduced complexity and temporal variability compared to the dynamic monsoon periods.

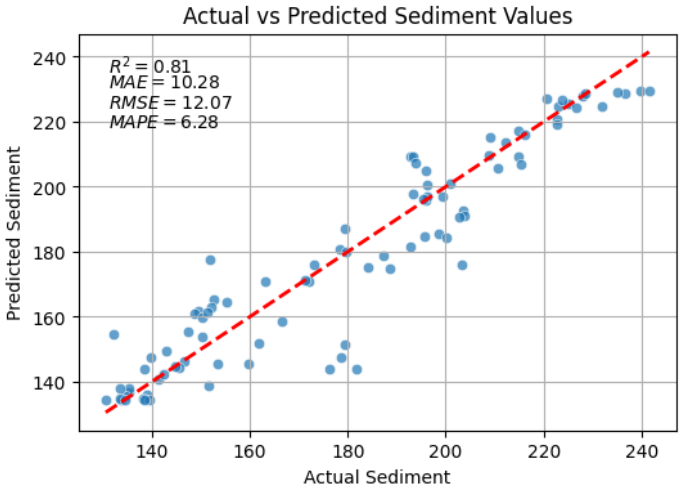

Figure 48: Scatterplot of SVM model -Winter Season

i) RF Pre-monsoon

The Random Forest (RF) regression model for pre-monsoon sediment prediction at Station 430.5 demonstrates moderate predictive performance, showing slight improvement over the KNN model for the same season but still facing challenges with the complex transitional dynamics of the pre-monsoon period. The model achieves an R^2 value of 0.24, indicating that approximately 24% of the variance in sediment concentration is explained by the discharge-based predictions, which is comparable to the KNN performance (0.25) but still reflects the inherent difficulty in modeling pre-monsoon sediment transport. The performance metrics reveal MAE of 64.28 ppm, RMSE of 87.35 ppm, and MAPE of 24.48%, showing reasonable prediction accuracy for moderate sediment concentrations but with notable errors for extreme values. The regression plot demonstrates that predicted values (red dots) generally follow the increasing trend of actual values (blue dots) across the discharge spectrum, but the model struggles significantly with higher discharge conditions (>22 m³/s)

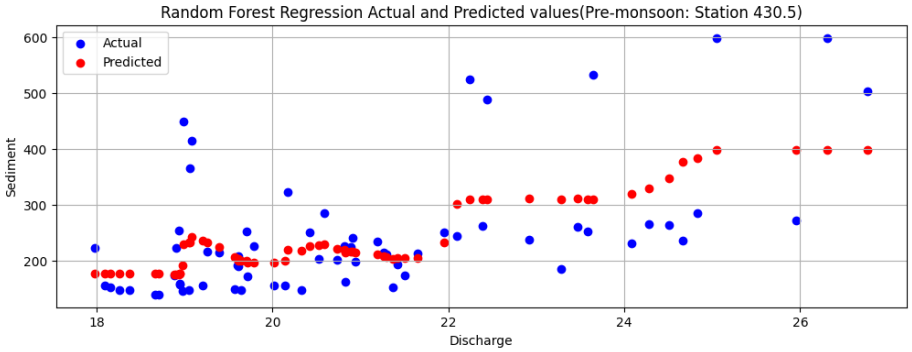

Figure 49: RF model for Pre-monsoon Season

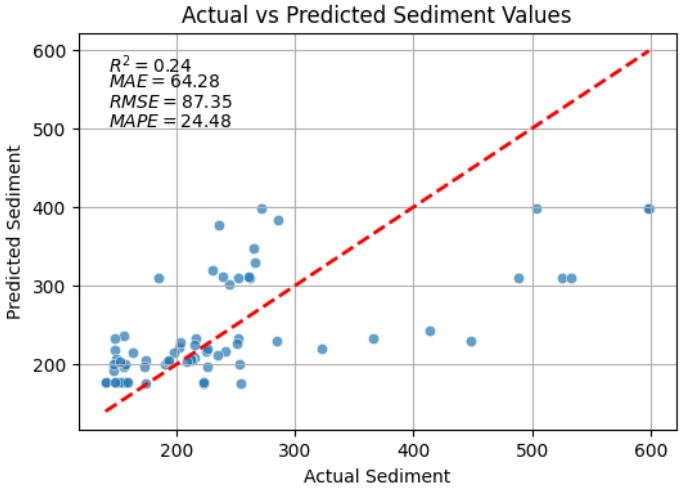

Figure 50: Scatterplot of RF model -Pre-monsoon Season

where it systematically underestimates peak sediment concentrations exceeding 400 ppm. The actual vs. predicted plot reveals considerable scatter around the 1:1 line, with predictions constrained to a narrower range (approximately 180-400 ppm) compared to actual values spanning 150-600 ppm, indicating that despite the ensemble approach's theoretical advantages in capturing non-linear relationships, the RF model cannot fully represent the complex snowmelt-driven and early rainfall dynamics that characterize pre-monsoon sediment mobilization.

j) RF Monsoon

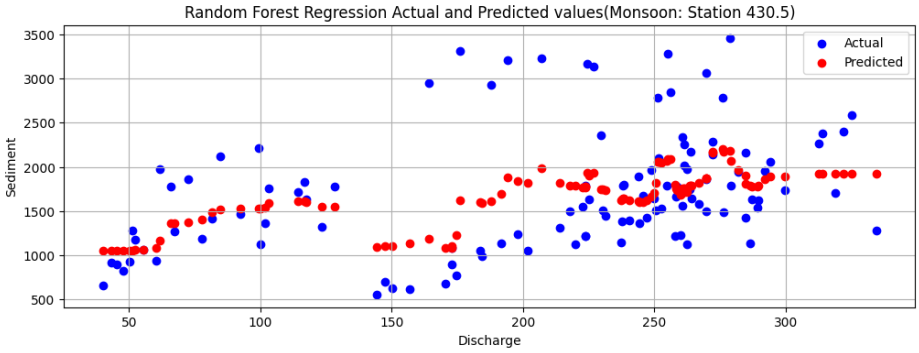

Figure 51: RF model for Monsoon Season

The Random Forest (RF) regression model exhibits poor predictive performance for monsoon sediment concentration at Station 430.5, demonstrating limited capability to handle the extreme variability and complex dynamics characteristic of the monsoon period. The model achieves a low R^2 value of 0.20, indicating that only 20% of the variance in sediment concentration is captured by discharge-based predictions, representing performance similar to the KNN model (0.22) but still inadequate for practical applications. The performance metrics reveal substantial prediction errors with MAE of 431.47 ppm, RMSE of 548.75 ppm, and MAPE of 23.22%, reflecting the model's inability to accurately predict the wide range of sediment concentrations encountered during monsoon conditions. The regression plot illustrates the fundamental limitation of the RF approach in handling monsoon complexity, with predicted values (red dots) forming a relatively constrained band that fails to respond adequately to the extreme variability in actual sediment concentrations (blue dots), particularly the dramatic peaks exceeding 3,000 ppm that dominate monsoon sediment transport. The actual vs. predicted plot demonstrates severe systematic bias, showing that while actual sediment concentrations span from 500 to over 3,500 ppm, RF predictions are artificially limited to approximately 1,000-2,300 ppm range, creating consistent underestimation of high sediment loads that are critical for annual sediment budgets. This poor performance indicates that even the sophisticated ensemble learning approach of Random Forest cannot adequately capture the multi-factorial complexity of monsoon sediment transport, where rainfall intensity, antecedent moisture conditions, sediment availability, timing effects, and hysteresis phenomena interact in ways that exceed the model's representational capacity.

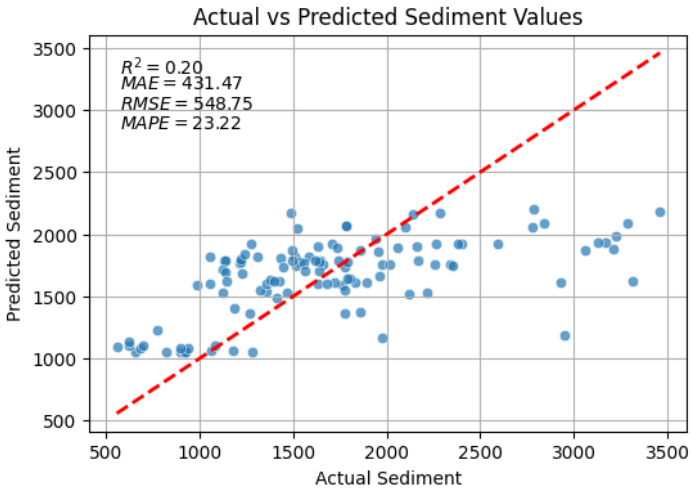

Figure 52: Scatterplot of RF model -Monsoon Season

k) RF Post-monsoon

The Random Forest (RF) regression model demonstrates outstanding predictive performance for post-monsoon sediment concentration at Station 430.5, achieving results that are virtually identical to the excellent

KNN performance for the same season. The model attains an exceptional R^2 value of 0.97, indicating that 97% of the variance in sediment concentration is successfully captured by the discharge-based predictions, with performance metrics showing remarkable accuracy: MAE of 12.95 ppm, RMSE of 17.22 ppm, and MAPE of 3.50%. The regression plot reveals that predicted values (red dots) align almost perfectly with actual values (blue dots) across the entire discharge spectrum from 40 to 140 m^3/s, demonstrating the ensemble model's ability to capture the systematic post-monsoon recession patterns with minimal deviation, while the actual vs. predicted plot shows data points clustering extremely tightly around the 1:1 line with virtually no scatter. This superior performance reflects the RF model's effectiveness in capturing the predictable, exponential decay patterns characteristic of post-monsoon sediment transport, where both discharge and sediment concentration follow consistent recession curves that create stable relationships ideally suited to the ensemble learning approach. The model's exceptional accuracy during this period demonstrates that Random Forest algorithms excel when applied to hydrological regimes with clear, systematic temporal patterns, making post-monsoon conditions particularly well-matched to the ensemble method's capacity for capturing consistent discharge-sediment relationships through multiple decision trees.

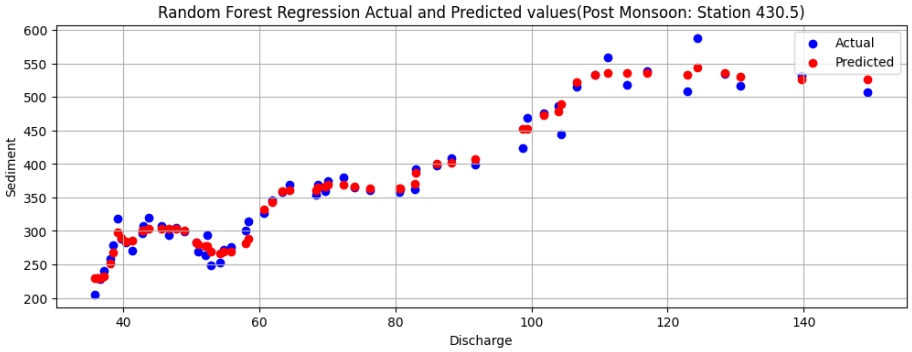

Figure 53: RF model for Post-monsoon Season

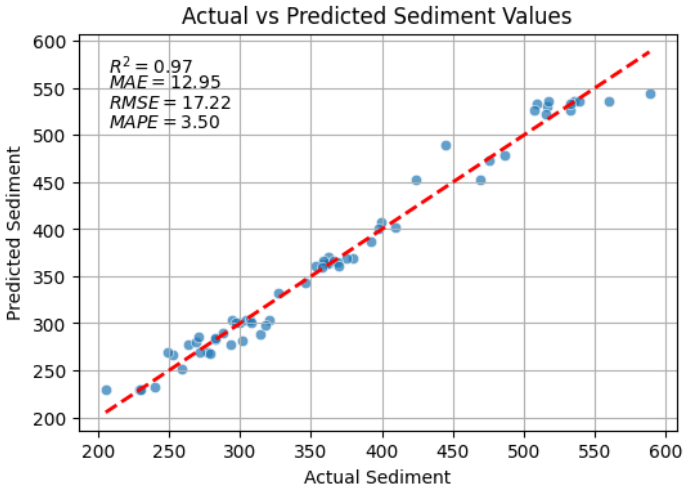

Figure 54: Scatterplot of RF model -Post-monsoon Season

l) RF Winter

The Random Forest (RF) regression model shows good predictive performance for winter sediment concentration at Station 430.5, demonstrating consistent accuracy across the stable, low-flow conditions that characterize the dry season. The model achieves a solid R^2 value of 0.70, indicating that 70% of the variance in sediment

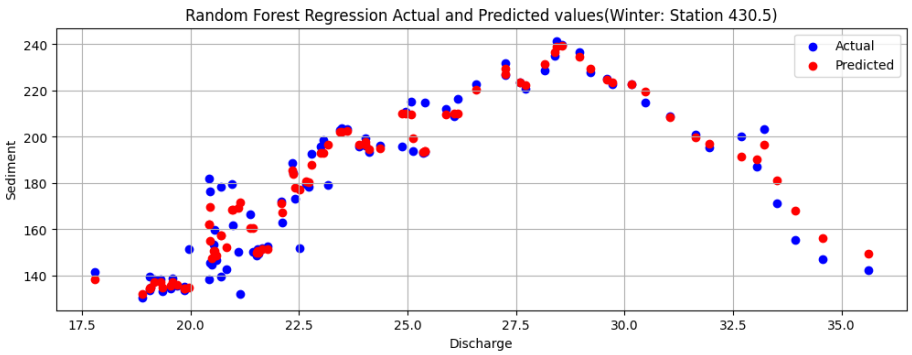

Figure 55: RF model for winter Season

concentration is explained by discharge-based predictions, with performance metrics showing reasonable accuracy: MAE of 11.88 ppm, RMSE of 15.19 ppm, and MAPE of 7.35%. The regression plot reveals that predicted values (red dots) closely follow actual values (blue dots) across the discharge range from 17.5 to 35 m³/s, with particularly strong agreement in the central discharge range where most winter data cluster around 20-25 m³/s, while the actual vs. predicted plot shows data points well-aligned along the 1:1 line with some scatter primarily at the lower and upper extremes of the sediment concentration range. This strong performance reflects the RF model's ability to capture the stable baseflow-dominated conditions during winter, where limited variability in both discharge and sediment concentration creates predictable patterns that the ensemble approach can model effectively through its multiple decision tree framework. The winter season's consistent hydrological behavior and minimal extreme events provide favorable conditions for Random Forest regression, demonstrating that ensemble methods perform well when applied to periods with reduced complexity and systematic relationships, though the performance is slightly lower than the post-monsoon period due to the inherently lower signal-to-noise ratio in winter's minimal sediment transport conditions.

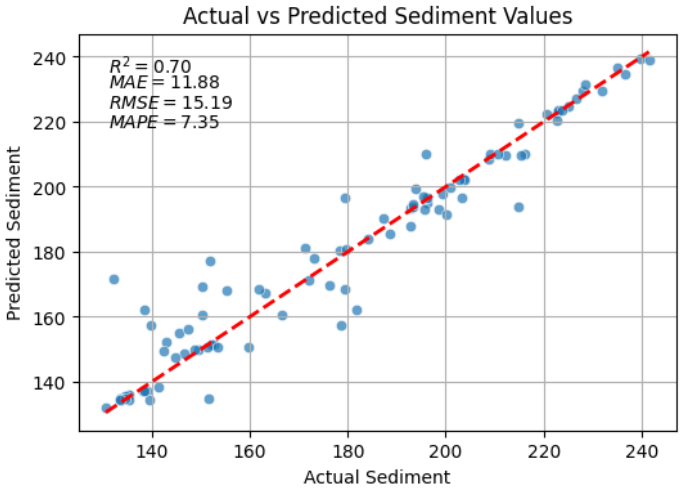

Figure 56: Scatterplot of RF model -Winter Season\

3.5 Model Performance Evaluation and Selection

Based on the comparative performance metrics presented in the table, the selection of optimal machine learning models varies significantly across seasons, reflecting the diverse hydrological conditions and sediment transport dynamics at Station 430.5. For the pre-monsoon season, SVM emerges as the preferred model with an R^2 of 0.25 and MAPE of 21.87%, outperforming both SVM (R^2 = 0.09, MAPE = 19.43%) and Random Forest (R^2 = 0.24, MAPE = 24.48%), though all models show limited predictive capability due to the complex transitional dynamics of snowmelt and early rainfall. The monsoon season presents the greatest modeling challenge, with KNN achieving the best performance (R^2 = 0.22, MAPE = 22.01%) compared to SVM's poor results (R^2 = -0.144, MAPE = 27.09%) and Random Forest's similar limitations (R^2 = 0.2045, MAPE = 23.22%), highlighting the extreme difficulty in capturing the non-linear, highly variable sediment transport during intense rainfall periods. In contrast, both post-monsoon and winter seasons demonstrate excellent model performance, with KNN and Random Forest achieving outstanding results (R^2 > 0.93 for post-monsoon, R^2 > 0.70 for winter), while SVM shows strong performance in these stable periods (R^2 = 0.9286 for post-monsoon, R^2 = 0.8087 for winter), indicating that the predictable recession and baseflow conditions during these seasons are well-suited to all machine learning approaches, making model selection less critical during these periods of consistent discharge-sediment relationships.

Table 3: Model Performance Evaluation

Station 430.5 (Seti Gandaki)						
Seasons	Models Used					
	KNN		SVM		Random Forest	
	R^2	MAPE	R^2	MAPE	R^2	MAPE
Pre-Monsoon	0.25	21.87%	**0.09**	**19.43%**	0.24	24.48%
Monsoon	**0.22**	**22.01%**	-0.144	27.09%	0.2045	23.22%
Post Monsoon	**0.97**	**3.35%**	0.9286	4.99%	0.9673	3.50%
Winter	0.74	6.79%	**0.8087**	**6.28%**	0.6970	7.35%

3.6 Development of Sediment Rating Curve

The sediment rating curves (SRCs) for Station 430.5 on the Seti-Gandaki River were developed using power-law relationships between discharge and sediment concentration, tailored to each of Nepal's four distinct hydrological seasons. These curves provide essential tools for estimating suspended sediment loads across varying flow conditions in the Gandaki River Basin, where continuous sediment monitoring faces significant logistical and financial constraints (Tfwala and Wang, 2016). During the **pre-monsoon season (Figure 57)**, the fitted power law equation is $S = 0.85 \times Q^{1.81}$, with a Mean Absolute Error (MAE) of 68.53 ppm and a Mean Absolute Percentage Error (MAPE) of 26.82%, reflecting the transitional dynamics of snowmelt-driven sediment mobilization and early rainfall contributions characteristic of this period.

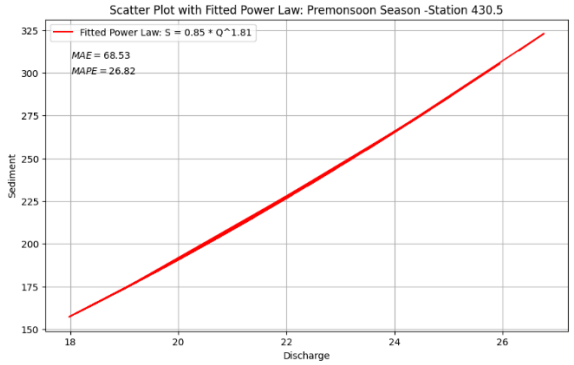

Figure 57: SRC for Pre-monsoon Season

The **monsoon season (Figure 58)** presents the most challenging conditions for traditional SRC development, with the power law relationship expressed as S = 333.66 × Q^0.31, achieving a MAE of 1564.66 ppm and a substantially elevated MAPE of 90.67%. This high error magnitude reflects the inherent complexity of capturing extreme sediment concentrations during intense monsoon flows, where non-linear sediment supply dynamics, hysteresis effects, and variable sediment availability create conditions that exceed the representational capacity of simple power-law relationships (Horowitz, 2003). The relatively low exponent (0.31) indicates a reduced sensitivity of sediment concentration to discharge changes during monsoon periods, suggesting that factors beyond simple flow magnitude including rainfall intensity, antecedent moisture conditions, and sediment supply limitations dominate sediment transport processes.

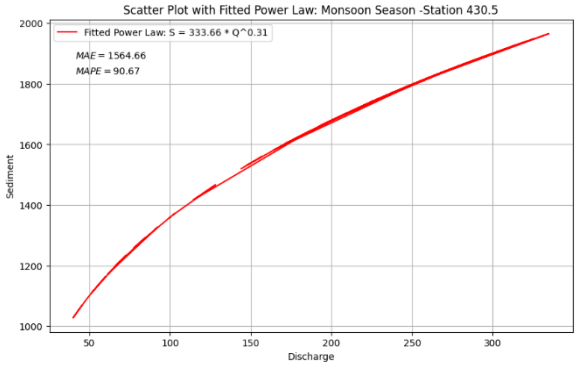

Figure 58: SRC for Monsoon Season

In contrast, the **post-monsoon season (Figure 59)** demonstrates significantly improved predictive accuracy with the equation $S = 26.55 \times Q^{0.62}$, yielding a substantially lower MAE of 26.44 ppm and MAPE of 7.66%. This marked improvement reflects the more stable and predictable recession patterns that characterize the post-monsoon period, where both discharge and sediment concentration follow systematic exponential decay curves that align well with traditional power-law modeling approaches (Horowitz, 2003). The **winter season (Figure 60)** exhibits the most stable sediment transport regime, with the power law relationship $S = 17.61 \times Q^{0.72}$ producing a MAE of 21.23 ppm and MAPE of 11.70%, indicating excellent model performance during baseflow-dominated conditions were minimal variability and consistent discharge-sediment relationships prevail.

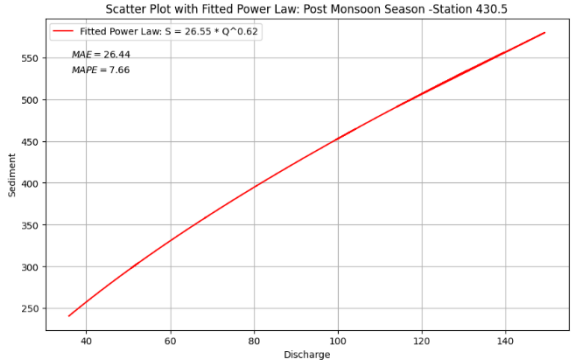

Figure 59: SRC for Post-monsoon Season

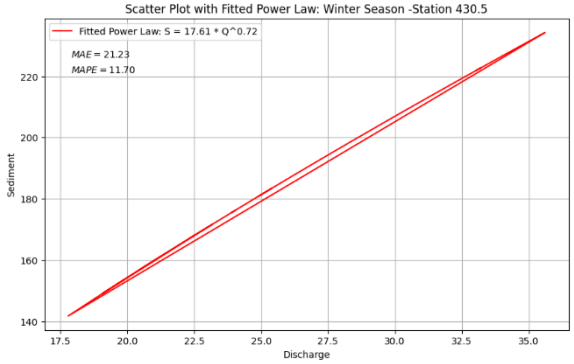

Figure 60: SRC for Winter Season

These seasonal-specific SRCs demonstrate the critical importance of considering hydrological variability when modeling sediment dynamics in monsoon-influenced river systems, ensuring more accurate predictions across diverse flow regimes in the Seti-Gandaki River. The development of season-specific sediment rating curves in this study underscores their essential role in estimating suspended sediment loads in data-scarce Himalayan River systems where continuous sediment monitoring remains challenging due to remote locations and resource limitations. SRCs provide practical and widely applicable tools for water resource management, particularly in regions where hydrological complexity necessitates sophisticated approaches to capture seasonal variations in sediment transport processes (Walling, Webb and Webb, 1988).

The varying exponents and coefficients across seasons reflect the influence of distinct hydrological processes: pre-monsoon snowmelt contributions (exponent = 1.81), monsoon-driven erosion and supply limitations (exponent = 0.31), post-monsoon stabilization and recession (exponent = 0.62), and winter baseflow conditions (exponent = 0.72). Notably, the extremely high MAPE of 90.67% during the monsoon season, while indicating substantial prediction challenges, aligns with findings from similar monsoon-influenced systems where traditional SRCs struggle to capture extreme sediment pulses driven by intense rainfall and complex sediment supply dynamics. This performance limitation supports the argument that while traditional SRCs remain valuable tools for stable flow periods, they require supplementation with advanced modeling approaches such as machine learning techniques or physically-based models to adequately represent the complex, time-dependent, and non-linear interactions that characterize sediment transport during extreme hydrological events in Himalayan River systems.

4 Conclusion

This research successfully established season-specific sediment rating curves (SRCs) for Station 430.5 in the Seti-Gandaki River through the application of optimized machine learning algorithms that capture the complex discharge-sediment relationships across Nepal's distinct hydrological seasons. The comprehensive analysis demonstrated pronounced seasonal variability, with monsoon periods exhibiting peak flows exceeding 1,400 m^3/s and sediment concentrations reaching over 20,000 ppm, contrasting sharply with winter baseflow conditions characterized by flows below 35 m^3/s and sediment concentrations under 250 ppm. Model selection was based on rigorous performance evaluation using coefficient of determination (R^2) and Mean Absolute Percentage Error (MAPE) metrics, revealing that K-Nearest Neighbors (KNN) provided optimal accuracy for monsoon (R^2 = 0.22) and post-monsoon (R^2 = 0.97) seasons, while Support Vector Machine (SVM) demonstrated superior performance during pre-monsoon (R^2 = 0.09) and winter (R^2 = 0.81) periods. The resulting power-law equations: $S = 0.85 \times Q^{1.81}$ (pre-monsoon), $S = 333.66 \times Q^{0.31}$ (monsoon), $S = 26.55 \times Q^{0.62}$ (post-monsoon), and $S = 17.61 \times Q^{0.72}$ (winter) reflect the distinct seasonal sediment transport mechanisms and provide robust frameworks for suspended sediment load estimation in data-limited environments.

While the machine learning approach significantly advanced traditional SRC development, the substantial prediction errors during monsoon conditions (MAPE = 90.67%) underscore the persistent challenges in modeling extreme, non-linear sediment transport processes characterized by complex hysteresis effects, variable sediment availability, and multi-factorial controls beyond simple discharge relationships. As demonstrated by (Tfwala and Wang, 2016), these limitations highlight the potential advantages of deep learning architectures, particularly Artificial Neural Networks with recurrent capabilities, which can capture temporal dependencies, antecedent conditions, and hierarchical feature interactions that conventional machine learning models cannot adequately represent. Consequently, while this study advances understanding of seasonal sediment dynamics and demonstrates the benefits of optimized model selection for SRC development, future research should prioritize the integration of advanced deep learning techniques,

additional hydro-meteorological variables, and longer temporal datasets to enhance predictive accuracy across all seasonal conditions. Such methodological advancements are crucial for supporting sustainable water resource management, hydropower infrastructure design, and climate adaptation strategies in the rapidly evolving Himalayan River systems where accurate sediment prediction remains critical for environmental and economic sustainability.

6 Acknowledgements

The authors express sincere gratitude to the Research Training and Consultancy Division (RTCD) of Kantipur Engineering College for providing essential funding, resources, and technical guidance that enabled the successful completion of this study. We acknowledge the division's unwavering commitment to promoting research excellence and innovation within the academic community. We extend our appreciation to the Department of Hydrology and Meteorology (DHM), Nepal, for generously providing the hydrological and sediment datasets that formed the foundation of this research. The authors are also thankful to colleagues and mentors whose constructive feedback, insights, and continuous encouragement proved invaluable throughout the research process.

References

Abid Almubaidin, M.A. *et al.* (2023) "Enhancing sediment transport predictions through machine learning-based multi-scenario regression models," *Results in Engineering*, 20, p. 101585. Available at: https://doi.org/10.1016/J.RINENG.2023.101585.

Aggarwal, C.C. (2017) *Outlier Analysis*. Second. New York, USA: Springer International Publishing. Available at: https://doi.org/DOI10.1007/978-3-319-47578-3.

Andermann, C. *et al.* (2012) "Sediment transfer and the hydrological cycle of Himalayan rivers in Nepal," *Comptes Rendus Geoscience*, 344(11–12), pp. 627–635. Available at: https://doi.org/10.1016/J.CRTE.2012.10.009.

Asselman, N.E.M. (2000) "Fitting and interpretation of sediment rating curves," *Journal of Hydrology*, 234(3–4), pp. 228–248. Available at: https://doi.org/10.1016/S0022-1694(00)00253-5.

Azamathulla, H.M. *et al.* (2010) "Machine Learning Approach to Predict Sediment Load - A Case Study," *Clean - Soil, Air, Water*, 38(10), pp. 969–976. Available at: https://doi.org/10.1002/clen.201000068.

Bajracharya, D. *et al.* (2025) "Season Specific Sediment Rating Curve Development Using Machine Learning: A Case Study of the Mahakali River Basin, Nepal," *Journal of Earth Science & Climatic Change*, 16. Available at: https://doi.org/10.4172/2157-7617.1000883.

Banadkooki, F.B. *et al.* (2020) "Suspended sediment load prediction using artificial neural network and ant lion optimization algorithm," *Environmental Science and Pollution Research 2020 27:30*, 27(30), pp. 38094–38116. Available at: https://doi.org/10.1007/S11356-020-09876-W.

Baniya, M.B. *et al.* (2019) "Hydraulic Parameters for Sediment Transport and Prediction of Suspended Sediment for Kali Gandaki River Basin, Himalaya, Nepal," *Water 2019, Vol. 11, Page 1229*, 11(6), p. 1229. Available at: https://doi.org/10.3390/W11061229.

Boukhrissa, Z.A. *et al.* (2013) "Prediction of sediment load by sediment rating curve and neural network (ANN) in El Kebir catchment, Algeria Prediction of sediment load by sediment rating curve and neural

network (ANN) in El Kebir catchment Prediction of sediment load by sediment rating curve and neural network (ANN) in El Kebir catchment, Algeria," *Algeria. Journal of Earth System Science*, 122(5). Available at: https://doi.org/10.1007/s12040-013-0347-2ï.

Carbonneau, P.E., Lane, S.N. and Bergeron, N. (2006) "Feature based image processing methods applied to bathymetric measurements from airborne remote sensing in fluvial environments," *Earth Surface Processes and Landforms*, 31(11), pp. 1413–1423. Available at: https://doi.org/10.1002/ESP.1341;ISSUE:ISSUE:DOI.

Chhetri, A. *et al.* (2016) "Assessment of Sediment Load of Langtang River in Rasuwa District, Nepal," *Journal of Water Resource and Protection*, 8(1), pp. 84–92. Available at: https://doi.org/10.4236/JWARP.2016.81007.

Ezzaouini, M.A. *et al.* (2022) "Predicting Daily Suspended Sediment Load Using Machine Learning and NARX Hydro-Climatic Inputs in Semi-Arid Environment," *Water 2022, Vol. 14, Page 862*, 14(6), p. 862. Available at: https://doi.org/10.3390/W14060862.

Hastie, T., Tibshirani, R. and Friedman, J. (2009) *The Elements of Statistical Learning*. New York, NY: Springer New York (Springer Series in Statistics). Available at: https://doi.org/10.1007/978-0-387-84858-7.

Horowitz, A.J. (2003) "An evaluation of sediment rating curves for estimating suspended sediment concentrations for subsequent flux calculations," *Hydrological Processes*, 17(17), pp. 3387–3409. Available at: https://doi.org/10.1002/HYP.1299;PAGE:STRING:ARTICLE/CHAPTER.

Hyndman, R.J. and Koehler, A.B. (2006) "Another look at measures of forecast accuracy," *International Journal of Forecasting*, 22(4), pp. 679–688. Available at: https://doi.org/10.1016/J.IJFORECAST.2006.03.001.

Jamal Chachan, L. and Sulaiman Bahnam, B. (2022) "Models for Predicting River Suspended Sediment Load Using Machine Learning: A Survey," *Technium*, Vol. 4, No.10. Available at: www.techniumscience.com.

Katwal, K. *et al.* (2025) "Flood Impact Assessment, LSTM-Based Forecasting, and Revenue Loss Analysis of the Melamchi Water Supply Project," *International Journal of Data Science and Analysis 2025, Volume 11, Page 125*, 11(5), pp. 125–135. Available at: https://doi.org/10.11648/J.IJDSA.20251105.11.

Khosravi, K. *et al.* (2022) "Suspended sediment load modeling using advanced hybrid rotation forest based elastic network approach," *Journal of Hydrology*, 610, p. 127963. Available at: https://doi.org/10.1016/J.JHYDROL.2022.127963.

Morin, G.P. *et al.* (2018a) "Annual Sediment Transport Dynamics in the Narayani Basin, Central Nepal: Assessing the Impacts of Erosion Processes in the Annual Sediment Budget," *Journal of Geophysical Research: Earth Surface*, 123(10), pp. 2341–2376. Available at: https://doi.org/10.1029/2017JF004460;WGROUP:STRING:PUBLICATION.

Morin, G.P. *et al.* (2018b) "Annual Sediment Transport Dynamics in the Narayani Basin, Central Nepal: Assessing the Impacts of Erosion Processes in the Annual Sediment Budget," *Journal of Geophysical Research: Earth Surface*, 123(10), pp. 2341–2376. Available at: https://doi.org/10.1029/2017JF004460.

Nda, M. *et al.* (2023) "An Overview of Machine Learning Techniques for Sediment Prediction," *Engineering Proceedings 2023, Vol. 56, Page 204*, 56(1), p. 204. Available at: https://doi.org/10.3390/ASEC2023-16599.

Oeurng, C., Sauvage, S. and Sánchez-Pérez, J.M. (2011) "Assessment of hydrology, sediment and particulate organic carbon yield in a large agricultural catchment using the SWAT model," *Journal of Hydrology*, 401(3–4), pp. 145–153. Available at: https://doi.org/10.1016/J.JHYDROL.2011.02.017.

Oli, K.P. and Zomer, R. (2010) "Kailash Sacred Landscape Conservation Initiative: Target Area Delineation Report." Available at: https://doi.org/10.53055/ICIMOD.528.

Pant, R.R. *et al.* (2018) "Spatiotemporal variations of hydrogeochemistry and its controlling factors in the Gandaki River Basin, Central Himalaya Nepal," *Science of The Total Environment*, 622–623, pp. 770–782. Available at: https://doi.org/10.1016/J.SCITOTENV.2017.12.063.

Pant, R.R. *et al.* (2025) "Trace elements in fluvial sediments of the Gandaki River Basin, Central Himalaya, Nepal: distribution, sources, and risk assessment," *Journal of Soils and Sediments 2025 25:8*, 25(8), pp. 2463–2480. Available at: https://doi.org/10.1007/S11368-025-04091-X.

Ponce, V.M. (2023) *Is a sediment rating curve really curved?*, *San Diego State University, San Diego, California*. Available at: https://ponce.sdsu.edu/sediment_rating_curve.html (Accessed: July 12, 2025).

Prasad, A. *et al.* (2020) "Spatial and Temporal Variation of Fish Assemblages in Seti Gandaki River, Tanahu, Nepal," *Borneo Journal of Resource Science and Technology*, 10(2), pp. 93–104. Available at: https://doi.org/10.33736/BJRST.2048.2020.

Salarijazi, M. *et al.* (2024) "Development of Suspended Sediment Rating Curve Model by Statistical Classification of River Discharge Data (Case Study: Ghareh-Sou Coastal Watershed)," *Iranian Journal of Science and Technology, Transactions of Civil Engineering 2024 48:6*, 48(6), pp. 4663–4672. Available at: https://doi.org/10.1007/S40996-024-01369-X.

Sharma, T.P.P. *et al.* (2020) "Assimilation of Snowmelt Runoff Model (SRM) Using Satellite Remote Sensing Data in Budhi Gandaki River Basin, Nepal," *Remote Sensing 2020, Vol. 12, Page 1951*, 12(12), p. 1951. Available at: https://doi.org/10.3390/RS12121951.

Shrestha, A.B. and Aryal, R. (2011) "Climate change in Nepal and its impact on Himalayan glaciers," *Regional Environmental Change*, 11(SUPPL. 1), pp. 65–77. Available at: https://doi.org/10.1007/S10113-010-0174-9/METRICS.

Tfwala, S.S. and Wang, Y.M. (2016) "Estimating Sediment Discharge Using Sediment Rating Curves and Artificial Neural Networks in the Shiwen River, Taiwan," *Water 2016, Vol. 8, Page 53*, 8(2), p. 53. Available at: https://doi.org/10.3390/W8020053.

Van, C.P., Le, H. and Chin, L. Van (2023) "Estimation of daily suspended sediment concentration in the Ca River Basin using a sediment rating curve, multiple regression, and long short-term memory model," *Journal of Water and Climate Change*, 14(12), pp. 4356–4375. Available at: https://doi.org/10.2166/WCC.2023.229.

Walling, D.E., Webb, B.W. and Webb, &b W. (1988) "The reliability of rating curve estimates of suspended sediment yield: some further comments," in *Proceedings of the Porto Alegre Symposium*. IAHS Publ.

WECS (2011) *Water Resources of Nepal in the Context of Climate Change*. Kathmandu. Available at: www.wec.gov.np.