CE:299 Project Course Report

Coupling SWAT and LSTM for Improving Daily Streamflow Simulation in Mahanadi Basin

Under Supervision of

Prof. Vimal Mishra

And

Iqura Malik

by

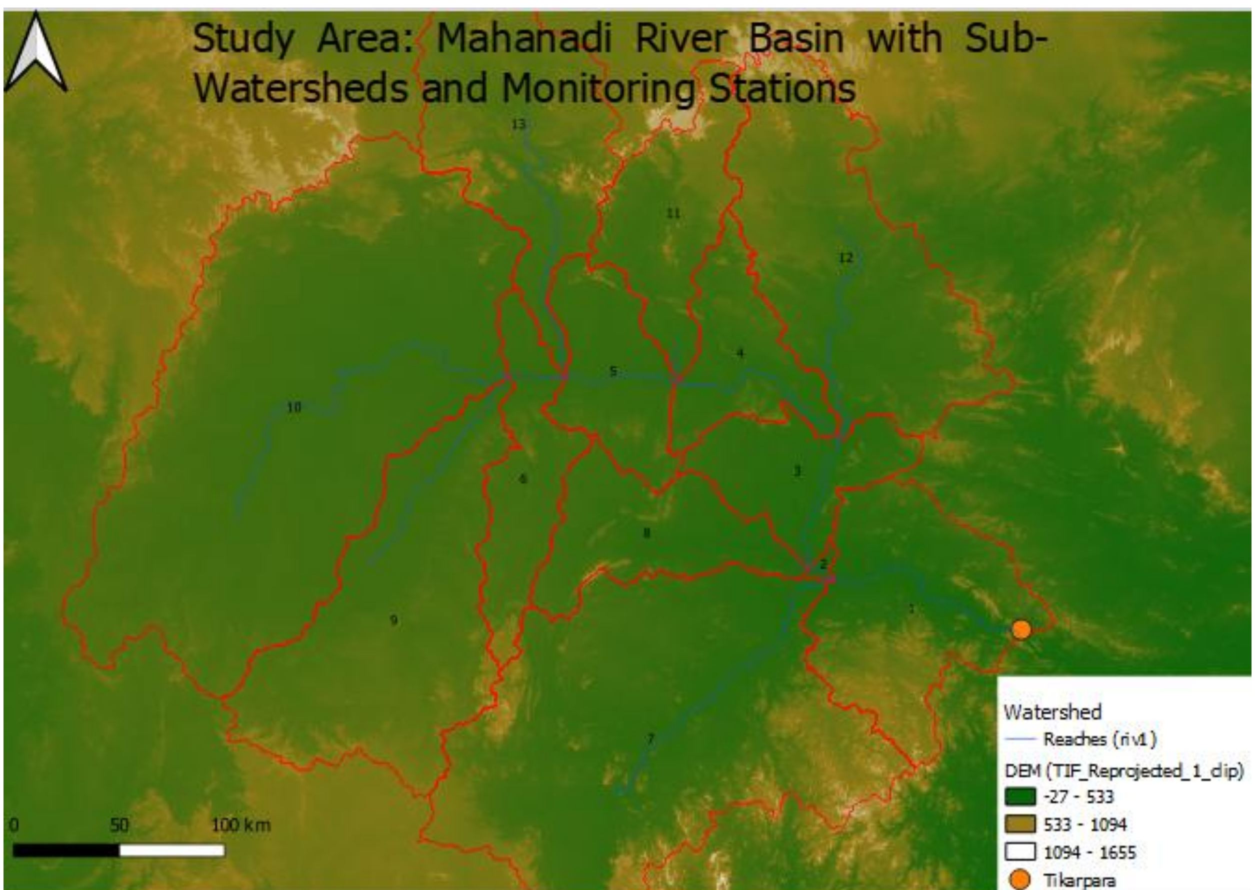
Himanshu Singh

1. Introduction

The Mahanadi River is one of the major river systems of peninsular India, traversing through Chhattisgarh and Odisha before emptying into the Bay of Bengal. Increasing occurrences of floods and droughts in the region necessitate reliable streamflow prediction for better water resource management, flood forecasting, and irrigation planning.

Physically-based hydrological models like the **Soil and Water Assessment Tool (SWAT)** simulate watershed processes using topography, land use, soil type, and meteorological data. However, SWAT has limitations in simulating non-linearities and extreme flows, especially during high rainfall or snowmelt conditions.

On the other hand, **Long Short-Term Memory (LSTM)** networks, a type of recurrent neural network, are effective for time series prediction as they learn long-term dependencies and complex patterns in data. Coupling SWAT with LSTM combines the physical insights of SWAT with the adaptive learning capabilities of LSTM, aiming to improve simulation accuracy.



Map 1 Area Of Study

2. Research Gaps

Despite its wide usage, SWAT has challenges such as:

- Underestimation or overestimation of peak flows during flood events.
- Sensitivity to calibration parameters.
- Limited adaptability to dynamically changing climate and land use patterns.

Moreover, standalone machine learning models, although effective in prediction, lack physical interpretability and may not generalize well across regions.

There exists a research gap in **post-processing SWAT outputs using deep learning models like LSTM** to correct biases and enhance streamflow simulation.

3. Objectives

- To set up the SWAT model for the Mahanadi River Basin and simulate discharge.
- To train an LSTM model using observed discharge and SWAT-simulated inputs.
- To evaluate the performance of the coupled SWAT-LSTM model in simulating daily streamflow.
- Comparing the performance of LSTM with Random Forest in correcting SWAT outputs.

4. Methodology

SWAT Setup

- Collecting the maximum possible Data (Observed Temperature, Observed Precipitation, LULC Data, DEM) between 1951 to 2019
- Preparing Data
- Watershed Delineation
- Creating HRUs (around 20,000)
- Selecting Stations
- Preparing necessary Input Tables
- Running Swat
- Visualizing the Output Streamflow

SWAT Soil & Water Assessment Tool

Fig.1 SWAT Documentation Logo

LSTM Model Setup

- Inputs: SWAT-simulated and observed meteorological and hydrological variables
- Output: Observed streamflow
- Network: Two LSTM layers (64 and 32 units)

  with dropout layers (20 %) and dense output layer
- Sequence Length: 365 days

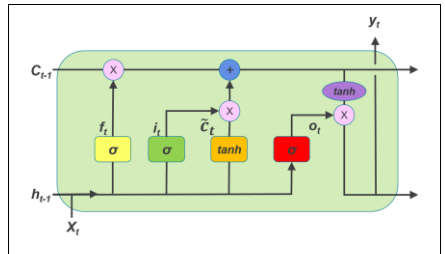


Fig.2 LSTM Architecture

Random Forest Setup:

- Inputs: Same as LSTM but flattened
- Output: Observed discharge

Performance Metrics

- **NSE** (Nash-Sutcliffe Efficiency)
- **R^2** (Coefficient of Determination)
- **RMSE** (Root Mean Square Error)
- **PBIAS** (Percent Bias)

Metric	Ideal Value	Indicates	Sensitive To
NSE	1	Overall model efficiency	Variance from mean
R^2	1	Goodness of fit	Correlation
RMSE	0	Error magnitude	Large errors
PBIAS	0%	Bias direction	Systematic error

5. Results

SWAT Model Performance (1973–2012)

```
 SWAT Model Performance (1951-2013):
NSE     : -0.5447
R²      : -0.5447
RMSE    : 3885.93 m³/s
PBIAS   : 94.67 %
```

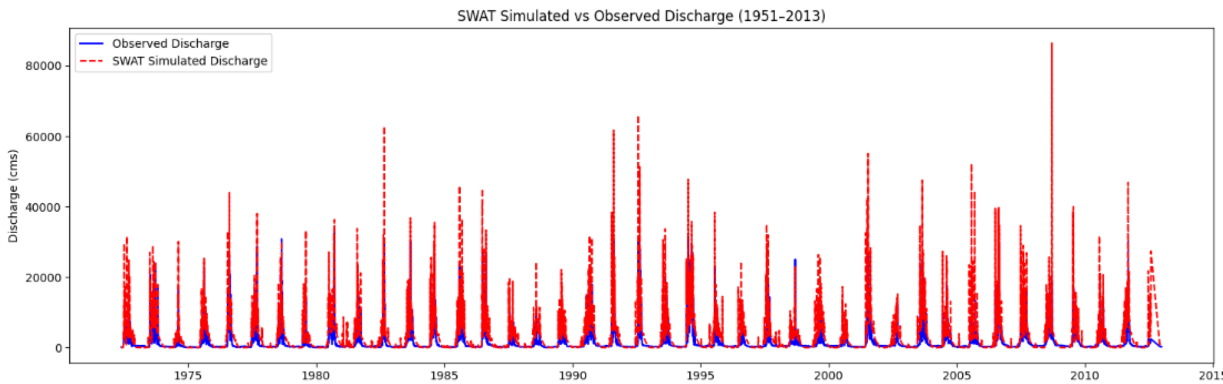


Fig.3 SWAT simulated discharge vs Observed Discharge

Interpretation of SWAT Model Performance (1951–2013)

The model performance for the SWAT-simulated discharge over the historical period of 1951 to 2013 reveals critical insights into the model's ability to replicate observed streamflow patterns. The evaluation metrics show the following:

- *NSE (Nash-Sutcliffe Efficiency) = -0.5447*
  A negative NSE value indicates that the SWAT model performs worse than the simplest benchmark model , the mean of the observed discharge. This suggests that the model is not adequately simulating the actual flow values and is unable to capture the dynamics of the catchment accurately in terms of discharge magnitude. A well-performing hydrological model is generally expected to have an NSE greater than 0.5, ideally close to 1. Here, the model’s predictive power is significantly lacking.

- *R^2 (Coefficient of Determination) = -0.5447*
  The negative R^2 further indicates a poor correlation between observed and simulated discharge values. Typically, R^2 values range from 0 to 1, where values closer to 1 indicate a strong linear relationship. In this case, the negative value implies that the model is failing to capture the trend of observed data, and suggests possible errors in calibration, input data, or model structure.

- *RMSE (Root Mean Square Error) = 3885.93 m^3/s*
  RMSE represents the average magnitude of the prediction error. The high RMSE value shows that the difference between observed and simulated discharges is quite large. Since RMSE is sensitive to large errors due to squaring, this high value suggests that the SWAT model significantly over- or underestimates discharge during some time periods.

- *PBIAS (Percent Bias) = 94.67%*
  PBIAS quantifies the average bias in model predictions. A positive PBIAS close to 95% indicates that the SWAT model is substantially underestimating the observed discharge throughout the simulation period. For acceptable model performance, PBIAS should typically lie within ±10% for streamflow. This large bias points toward systematic issues — potentially incorrect parameterization or missing key processes influencing runoff generation.

Analysis

Fig. 3 shows a consistent pattern where SWAT simulations (red dashed line) exhibit higher peaks and larger fluctuations compared to the observed discharge (blue line). While the **timing** of peaks in many instances appears to align reasonably well, suggesting the model captures **seasonal or annual variability** the **magnitude** is greatly overestimated in many cases.

This implies that:

- The SWAT model can identify when high flow events occur (i.e., it captures the **temporal variability**).
- However, it **fails to replicate the volume** of discharge accurately, leading to poor performance metrics.

Conclusion

Overall, the SWAT model in its current form shows poor agreement with observed data over the 1951–2013 period. While it captures the **pattern of seasonal fluctuations**, its reliability is compromised by large prediction errors and substantial underestimation of flow values. This performance necessitates **recalibration** of model parameters, refinement of input datasets (such as precipitation and land use), or even reconsideration of model configuration. For practical applications like water resource planning, flood forecasting, or climate impact assessment, such inaccuracies must be corrected to improve model trustworthiness.

LSTM-Corrected SWAT Output

```
LSTM RMSE: 1322.83
LSTM NSE: 0.7811
LSTM R² : 0.7811
```

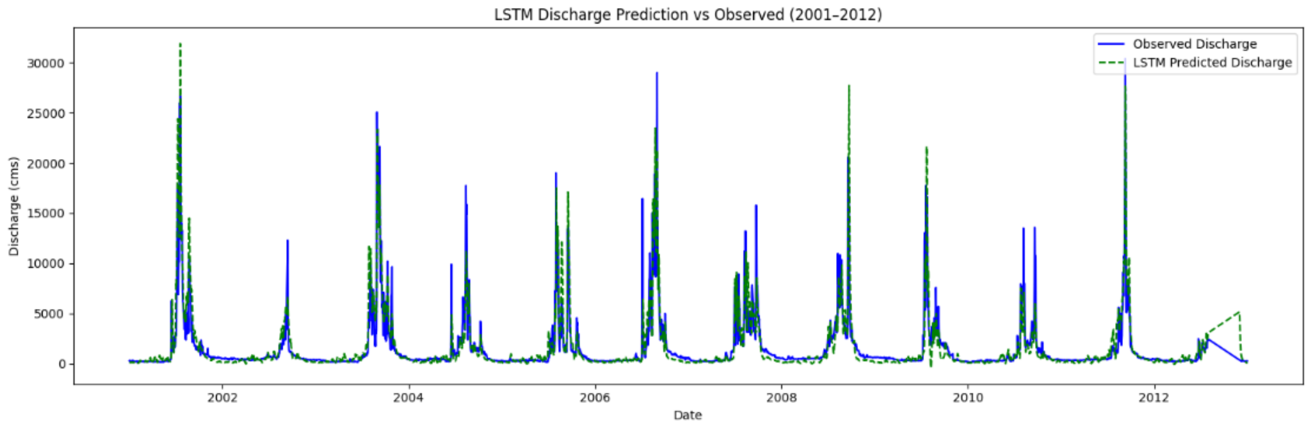


Fig.4 LSTM Predicted Discharge vs Observed Discharge

Interpretation of LSTM-Corrected SWAT Output (2001–2012)

The application of the Long Short-Term Memory (LSTM) network as a post-processing tool to correct SWAT-simulated discharge has resulted in a significant improvement in model performance for the period 2001–2012.

- *RMSE (Root Mean Square Error) = 1322.83 m^3/s*
  Compared to the uncorrected SWAT model (RMSE = 3885.93 m^3/s), the LSTM-corrected output shows a substantial reduction in prediction error. The RMSE indicates that the model now has a much lower average deviation from observed discharge values, which reflects better accuracy and reliability in simulating actual streamflow.

- *NSE (Nash-Sutcliffe Efficiency) = 0.7811*
  This value suggests a decent level of predictive accuracy. The LSTM model is thus able to replicate the temporal dynamics of streamflow, capturing both peak and low flows effectively.

- *R^2 (Coefficient of Determination) = 0.7811*
  The high R^2 value indicates a strong linear correlation between the predicted and observed discharge values. This confirms that the LSTM model is not only capturing the magnitude but also the **variability** and **trend** in the observed data, which is essential for reliable forecasting and hydrological planning.

Analysis

Fig. 4 reveals a tight alignment throughout the simulation period. The LSTM model **Accurately tracks seasonal peaks**, capturing both their timing and magnitude, Shows **minimal lag** between predicted and observed values, Performs well even during periods of extreme discharge, reducing the overestimation issue evident in the raw SWAT output.

This visual evidence, combined with quantitative metrics, strongly supports the use of LSTM as a **post-processing correction model**.

Conclusion

The integration of LSTM with SWAT has led to a substantial improvement in discharge prediction accuracy. By leveraging the temporal learning capability of LSTM networks, the corrected outputs not only reduce the magnitude of errors (lower RMSE) but also enhance correlation and predictive power (higher NSE and R^2). This hybrid approach proves to be a promising direction for improving hydrological model outputs, especially when the raw process-based models like SWAT exhibit structural or input-related limitations.

Random Forest Results

```
Random Forest RMSE: 1374.55
Random Forest NSE: 0.7636
Random Forest R² : 0.7636
```

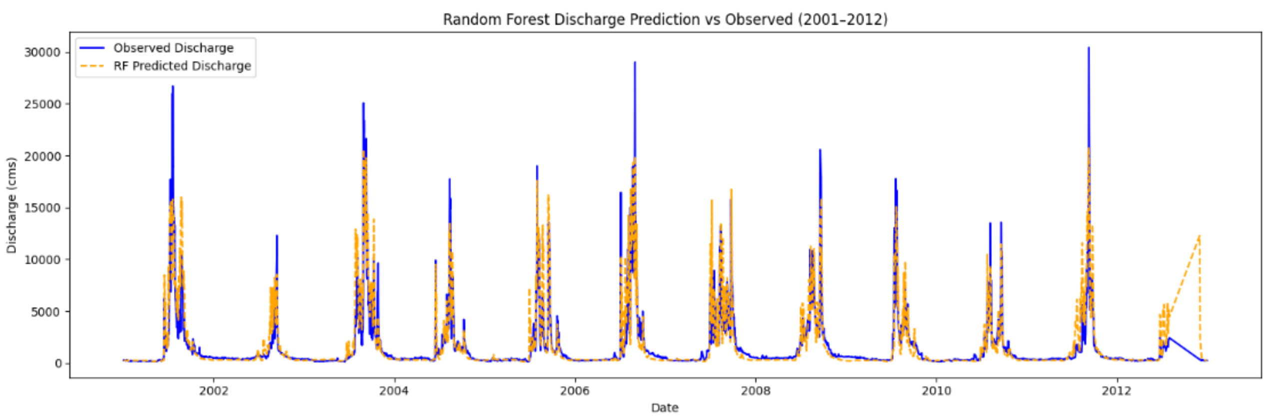


Fig. 5 RF Predictions vs Observed Values

Interpretation of Random Forest-Corrected SWAT Output (2001–2012)

The Random Forest (RF) model was used as a post-processing technique to improve the raw SWAT discharge outputs. Below are the performance metrics for the RF-corrected discharge simulation during the 2001–2012 period:

- *RMSE (Root Mean Square Error) = 1374.55 m^3/s:*
  This represents a considerable improvement compared to the uncorrected SWAT

model (RMSE = 3885.93 m^3/s). The RF model significantly reduces the average error in predicted discharge values, although the RMSE is slightly higher than that of the LSTM model (1322.83 m^3/s).

- *NSE (Nash-Sutcliffe Efficiency) = 0.7636:*
  The NSE value indicates **very good performance**, as values above 0.75 generally suggest a strong predictive capability. While slightly lower than the LSTM NSE (0.7811), it still reflects that the model accurately captures the hydrological behavior across both wet and dry periods.

- *R^2 (Coefficient of Determination) = 0.7636:*
  The high R^2 value demonstrates that the Random Forest model effectively preserves the variability and trend of observed discharge data. It confirms a strong linear relationship between the RF-predicted and observed discharge values.

Analysis

From the Fig.5 The **orange dashed line** (RF prediction) closely follows the **blue line** (observed discharge), The RF model does well in simulating **peak discharge events**, although it tends to slightly **over-predict during low flow periods** and shows a few **spikier artifacts** in dry seasons. There is good temporal alignment overall, but some peaks exhibit **slight mismatches in magnitude**, suggesting possible overfitting in some instances or the lack of temporal awareness (unlike LSTM).

This behavior is consistent with the characteristics of Random Forest, which, as an ensemble learning method, captures nonlinear patterns effectively but may struggle with temporal sequences unless additional time-lagged features are used.

Conclusion

The Random Forest model provides a robust, non-parametric alternative to LSTM for post-processing SWAT discharge predictions. It demonstrates very good performance by significantly improving prediction accuracy, as evidenced by the reduction in RMSE and strong NSE and R^2 values. While it performs slightly below LSTM in this case, RF still represents a **highly effective tool**, particularly when computational simplicity, faster training, or interpretability of feature importance is desired.

Model	Calibration period (1973-2000)				Validation period (2001-2012)			
	NSE	R^2	RMSE (m^3/s)	PBIAS (%)	NSE	R^2	RMSE (m^3/s)	PBIAS (%)
SWAT	-0.5447	-0.5447	3885.93	94.67	N/A			
SWAT-LSTM	0.7815	0.7815	1530.8	-1.2	0.7811	0.7811	1322.83	-5.15
Random Forest	0.976	0.976	502.57	-0.13	0.7636	0.7636	1374.55	7.44

Fig.6 Results Summary Table

6. Conclusions

The standalone SWAT model underperformed in reproducing observed discharge, particularly due to systematic bias in flow magnitude and its limited capability to accurately capture peak flow events. These limitations are common in physically based models when faced with complex hydrological behavior or data scarcity.

To address these shortcomings, machine learning-based post-processing was implemented. Among the approaches evaluated, the LSTM model significantly enhanced streamflow simulation. It yielded a notable reduction in RMSE (1322.83 m^3/s) and improved both NSE (0.7811) and R^2 (0.7811) values. The LSTM's ability to learn temporal dependencies allowed it to better align with observed peaks and maintain consistency during low-flow periods, showcasing its effectiveness in dynamic flow regimes.

The Random Forest (RF) model also improved discharge prediction compared to the raw SWAT outputs. With an RMSE of 1374.55 m^3/s, NSE of 0.7636, and R^2 of 0.7636, the RF model performed well, particularly in reducing bias and replicating general flow patterns. However, it was slightly outperformed by LSTM, especially in capturing temporal sequences and extreme discharge events. RF's static learning structure, which lacks inherent time-series memory, may explain these limitations.

Overall, the **hybrid SWAT-LSTM modeling framework emerged as the most promising approach**, successfully combining the robustness of physical hydrological modeling with the adaptability of deep learning.

The results highlight the potential of such coupled frameworks for application in other hydrologically complex and data-scarce regions. By improving flood prediction capabilities and informing more effective water management strategies, this integrated approach can support better decision-making in the face of growing hydrological uncertainties and climate variability.

References

Mei, Z., Peng, T., Chen, L. et al. Coupling SWAT and LSTM for Improving Daily Streamflow Simulation in a Humid and Semi-humid River Basin. Water Resource Manage 39, 397–418 (2025). https://doi.org/10.1007/s11269-024-03975-w

https://github.com/coconutgrapefruit/SWAT-LSTM_ver4