Project title: *Improving River Discharge Prediction in Data-Limited Regions using MANN-Transformer Architecture*

Presented By: Rashi Agrawal (22CH10053), Devika K V (22NA10007)

Supervisor / Mentor: *Ms. Somitra Sarkar*

---

1. Problem statement & study area

1.1 Problem statement

Accurate river discharge prediction is essential for water resource management, but many gauge stations suffer from limited historical data, making model training challenging. Traditional transfer learning approaches achieve good performance but often lack interpretability. To address this, we explore few-shot learning using a Memory-Augmented Neural Network (MANN) combined with transformer architectures. The objective is to develop a few-shot capable, reliable, and interpretable model that can perform well in data-scarce hydrological regions.

1.2 Study area

The study focuses on multiple gauging stations in the Mahanadi / Hirakud river system (the highlighted red dots in the map — see Figure 1). These stations span different parts of the basin (upper reach, middle reach, and reservoir-influenced reaches) and therefore sample contrasting hydrological regimes, catchment sizes, and anthropogenic controls. Testing the model across these stations allows us to evaluate robustness to geographic variability and varying lengths and quality of observational records.

Stations studied:

- **Region A — Basantpur:** mid-basin station with moderate catchment area and a strongly seasonal monsoon-driven hydrograph.
- **Region B — Hirakud:** downstream reservoir-influenced station affected by storage/release operations at Hirakud Reservoir (expected dampening of peak flows and altered seasonality).
- **Region C — Rajim:** tributary/mid-basin station with a smaller catchment, typically more flashy responses to rainfall.
- **Region D — Simga:** upper- or mid-basin station with limited historical records relative to the basin average.
- **Region E — Sundargarh:** headwater station representing a different sub-catchment and climatic influence; often shows fresher rainfall–runoff responses.
- **Region F — Kurubhata:** mid-to-upper basin station with distinct catchment characteristics; useful for testing transferability across catchments.

- **Region G — Ghatora:** tributary station that captures distinct hydrograph shapes and helps evaluate model generalization on smaller basins.

Temporal coverage & resolution:

2000-06-01 00:00:00 to 2014-09-30 00:00:00 (train:2005-06-01 00:00:00 to 2010-09-30 00:00:00, test: 2011-06-01 00:00:00 to 2014-09-30 00:00:00).
Daily discharge (Q, m^3/s), Mean-areal rainfall

Experimental sampling strategy:

To evaluate model performance under progressive scarcity, we ran experiments using four training fractions: **100%**, **80%**, **50%**, and **20%** of the available training records. By default, these subsets were produced by chronological truncation (keeping the earliest X% for training and reserving the remaining records for validation/test) to preserve temporal consistency.

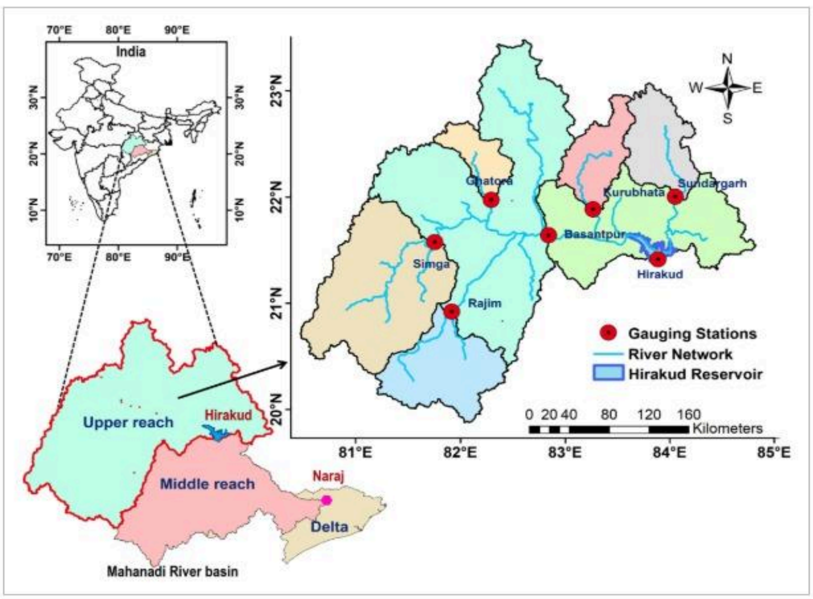


Figure 1. Study area and gauging stations. Map of the Mahanadi / Hirakud basin showing the gauging stations used in this study (highlighted red dots).

---

2. Method

2.1 Data

The experiments use daily time-series observations from multiple gauging stations in the Mahanadi / Hirakud basin. Each sample is assembled as a supervised example using a fixed look-back window of historical values (default **N_STEPS_IN = 30** days) and a short multi-step prediction horizon (default **N_STEPS_OUT = 5** days). Input feature sets include the target discharge at the station plus auxiliary covariates (e.g., mean areal rainfall, upstream station discharge) as available for the station of interest. For each station, we run experiments with varying training fractions to simulate data scarcity (typical fractions used **100%, 80%, 50%, 20%**).

2.2 Preprocessing

1. **Missing data handling.** Small gaps are imputed with forward/backward fill; larger gaps are excluded from training.
2. **Smoothing.** To reduce high-frequency noise and produce stable learning targets, a Blackman smoothing window is applied to both training and test series before model training. The typical Blackman window length used in the scripts is **N = 20**; the smoothed series are used for model fitting, but the final evaluation is aligned back to the raw, unfiltered test values using an offset procedure.
3. **Scaling.** Input features are scaled with **MinMaxScaler** (range 0–1), fit on the training split, and persisted for inverse transforms. The scaler is applied to both training and test sets.
4. **Feature engineering & sequence construction.** From the scaled multivariate array, we create supervised sequences using split_sequences(sequences, n_steps_in, n_steps_out, target_idx). This function slides a window of length **N_STEPS_IN** to create X (shape: samples × n_steps_in × n_features) and y (shape: samples × n_steps_out). Typical training/validation split inside the training fraction is 80/20.

2.3 Model(s)

The principal architecture implemented is an **MLP + Memory-Augmented Neural Network (MANN)** model:

- **Input projection.** The input sequence (flattened) is projected into a lower-dimensional embedding (memory_dim) via a dense layer.
- **Memory module.** A learnable memory matrix (memory_size × memory_dim, default **memory_size = 128**, **memory_dim = 64**) is stored as a parameter. The model computes attention scores between the projected input and each memory slot, obtains attention weights via softmax, and produces a memory read vector as the weighted sum of memory rows.
- **Augmented MLP head.** The flattened input and memory read vector are concatenated and passed through two dense layers (hidden_dim default 256 then 128) with ReLU activation, followed by a final linear layer that outputs the **N_STEPS_OUT** predictions.

Baseline models implemented for comparison include a plain MLP (no memory) and a Transformer encoder without an external memory.

2.4 Training procedure

- **Loss & optimizer.** Models are trained with **MSELoss** and the **Adam** optimizer (typical learning rate **1e-4**). Gradient clipping (clip_grad_norm_max_norm = 1.0) is applied to stabilize training.
- **Hyperparameters.** Default values used: **BATCH_SIZE = 32**, **EPOCHS = 50**, **LEARNING_RATE = 1e-4**, with reproducible seeds set using np.random.seed(42) and torch.manual_seed(42).
- **Validation & checkpointing.** A validation split monitors performance each epoch. The checkpoint with the lowest validation loss is saved. Early stopping can be enabled or emulated by monitoring validation loss across epochs and breaking training if no improvement is seen for a configurable patience.
- **Offset alignment to raw data.** Because the model trains on smoothed targets, the code implements a post-processing offset step to align smoothed predictions with the raw observed series: load the raw, unfiltered test target file, compute offset = raw_test - smoothed_test, and add this offset to the smoothed model predictions before final metric calculations. This preserves the model's learned dynamics while ensuring fair comparison against the original measurements.
- **Repetition & reporting.** Each experimental condition (region × percent-of-data) is typically repeated with multiple random seeds; reported metrics are means (and optionally standard deviations) across repeats.

2.5 Evaluation metrics

Performance is computed per prediction horizon step (t+1 … t+5) and aggregated summaries are reported. Key metrics:

- **NSE (Nash–Sutcliffe Efficiency):** fraction of variance explained by the model (1 = perfect, ≤0 poor).
- **RSR:** RMSE divided by the standard deviation of observations; lower is better (0 = ideal).
- **PBIAS (%):** percentage bias = 100 * (sum(predicted − observed) / sum(observed)); shows systematic over/underestimation.
- **TPE (%) (Total Percentage Error):** sum of absolute errors divided by sum of observations, expressed as a percentage.
- **RMSE / MAE:** RMSE penalizes large errors (sensitive to outliers); MAE is the mean absolute error (average magnitude of errors).

Metrics are computed on the offset-adjusted predictions versus the raw tests

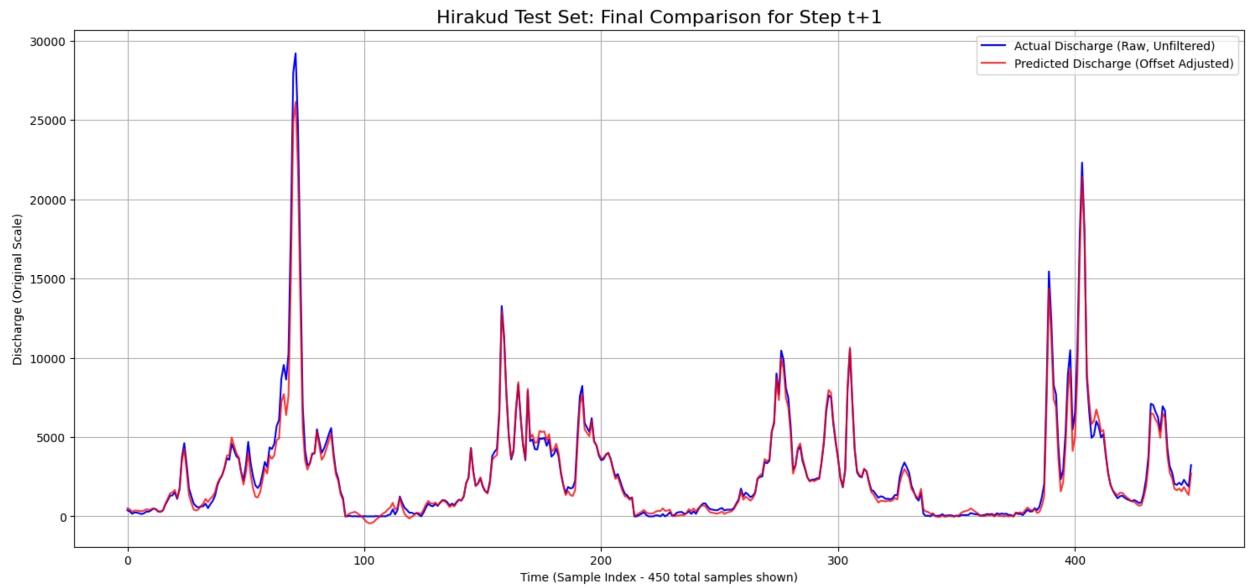
---

3. Results (for each region)

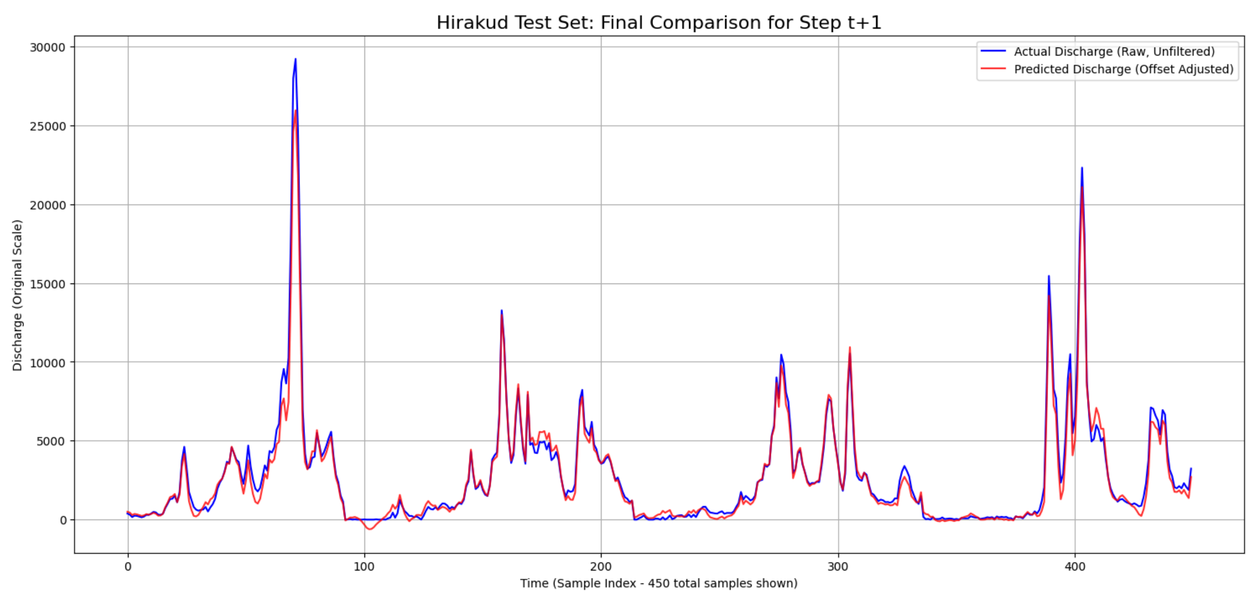
1. Hirakud

(a) Dataset Details & Plots

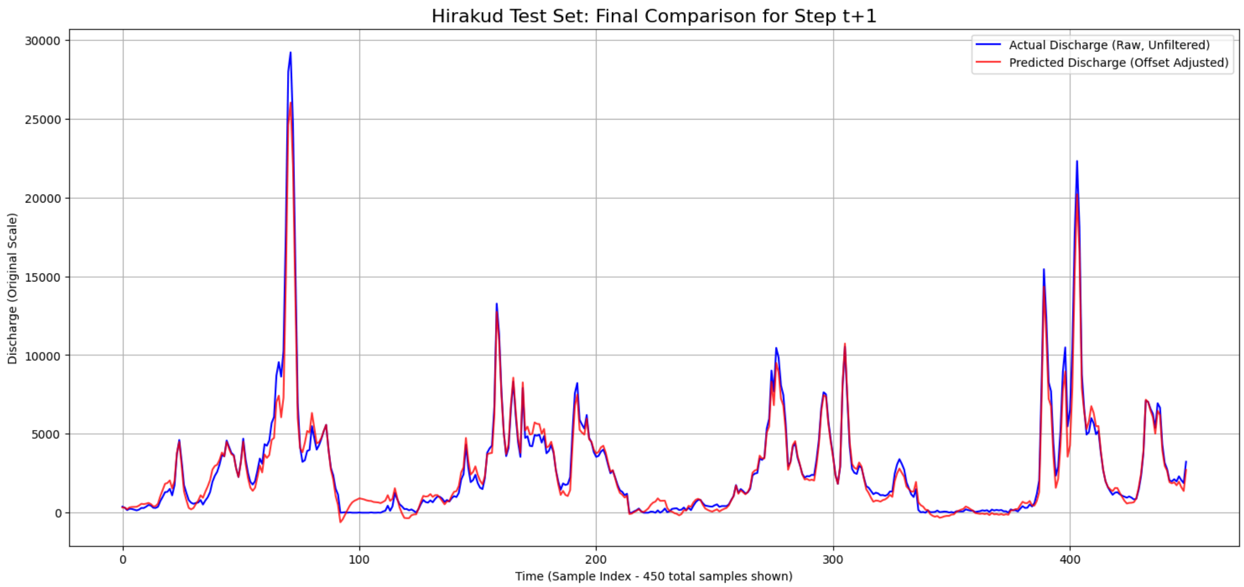
- **100% Data** (Train: [2000-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



- **80% Data** (Train: [2002-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



- **50% Data** (Train: [2005-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



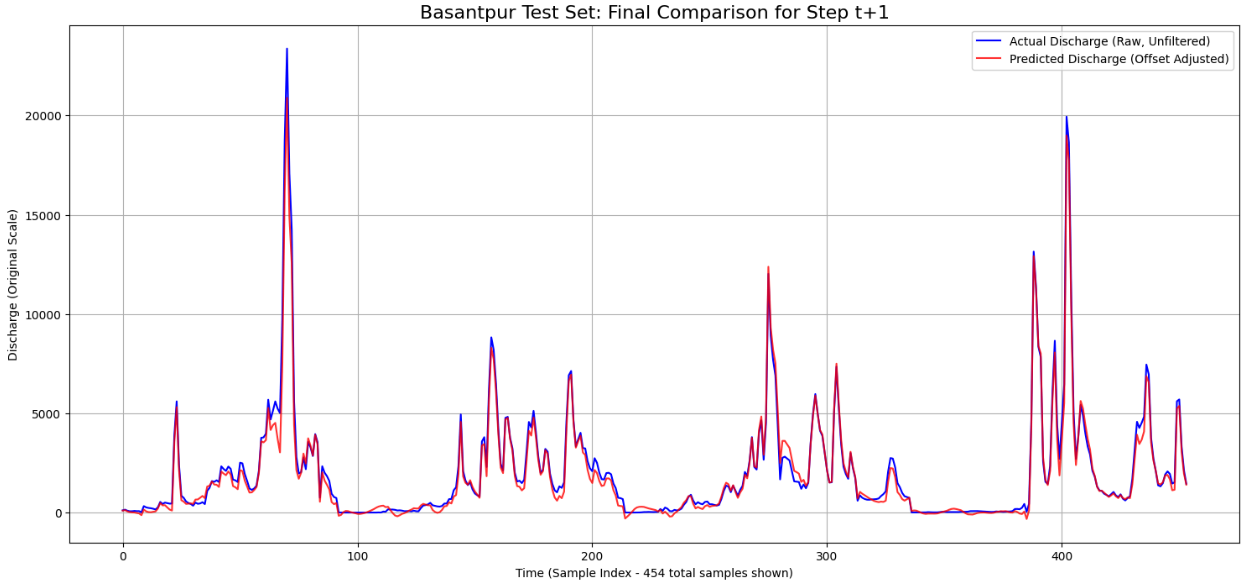
(b) Performance Metrics

Training %	NSE (t+1 → t+5)	RSR (t+1 → t+5)	PBIAS (%) (t+1 → t+5)	TPE (%) (t+1 → t+5)
100%	0.9821, 0.9695, 0.9515, 0.9168, 0.876	0.134, 0.1747, 0.2202, 0.2884, 0.3521	-4.9745, -6.4146, -6.7359, -8.028, -9.3794	10.3102, 13.8868, 18.1677, 23.0387, 27.845
80%	0.977, 0.9578, 0.9387, 0.8987, 0.8609	0.1518, 0.2053, 0.2475, 0.3184, 0.373	-6.0818, -8.0593, -8.105, -9.3906, -9.5173	12.411, 16.652, 20.5146, 25.8324, 30.0488
50%	0.972, 0.9576, 0.9328, 0.8992, 0.8593	0.1672, 0.2059, 0.2593, 0.3175, 0.3751	-3.4729, -3.3838, -4.3271, -5.108, -3.3216	14.2341, 17.495, 21.7368, 26.2727, 31.6282

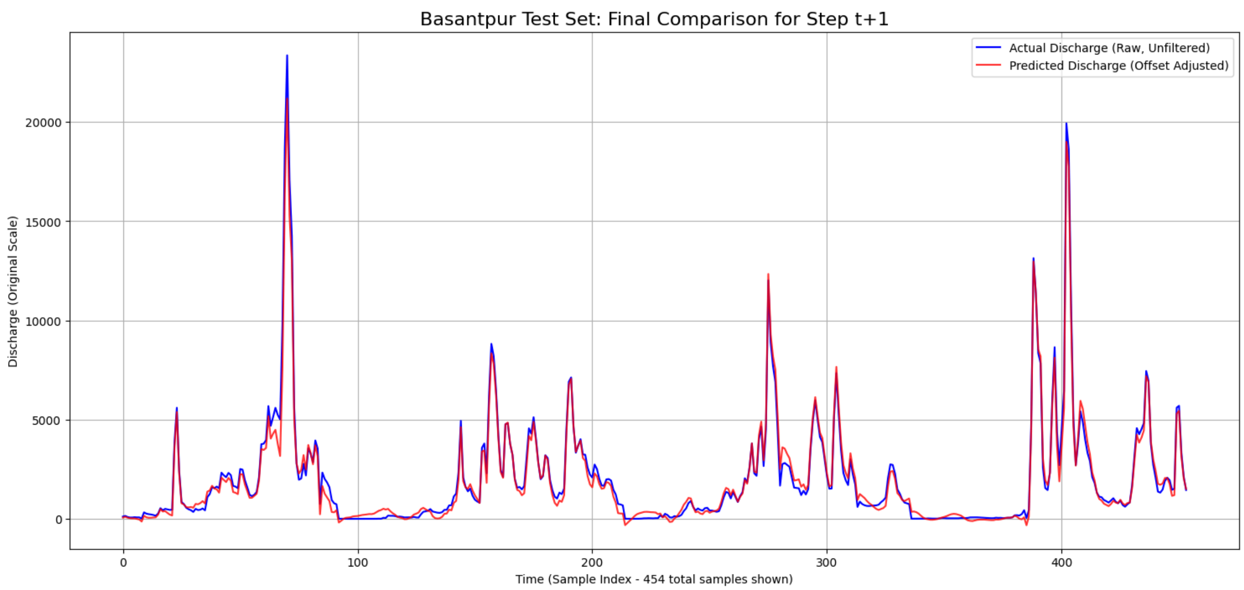
2. Basantpur

(a) Dataset Details & Plots

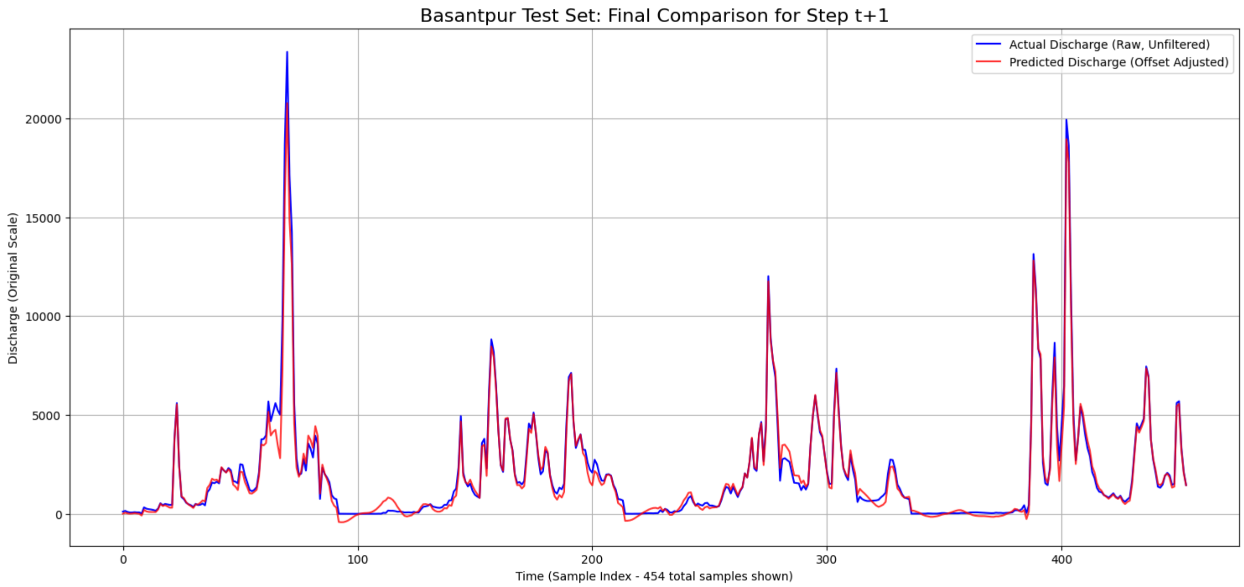
- **100% Data** (Train: [2000-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



- **80% Data** (Train: [2002-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



- **50% Data** (Train: [2005-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



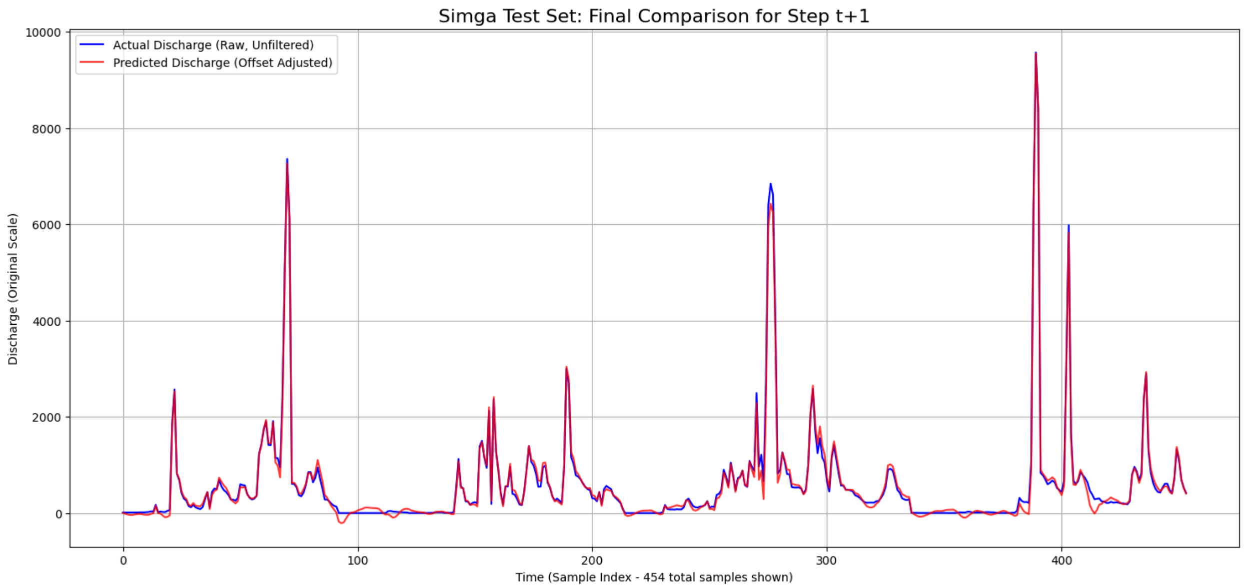
(b) Performance Metrics

Training %	NSE (t+1 → t+5)	RSR (t+1 → t+5)	PBIAS (%) (t+1 → t+5)	TPE (%) (t+1 → t+5)
100%	0.9812, 0.9634, 0.9332, 0.8936, 0.8447	0.137, 0.1914, 0.2585, 0.3262, 0.3941	-4.366, -5.1434, -7.0813, -9.4172, -9.6464	11.4484, 15.3585, 20.8195, 26.9643, 32.4589
80%	0.9812, 0.9679, 0.9458, 0.9108, 0.8596	0.137, 0.1791, 0.2327, 0.2987, 0.3747	-6.69, -7.6481, -7.936, -9.752, -11.3693	12.1699, 14.8973, 19.087, 24.2271, 30.4666
50%	0.9831, 0.9685, 0.9416, 0.9042, 0.8529	0.1301, 0.1774, 0.2417, 0.3096, 0.3836	-2.8289, -2.097, -4.3395, -6.2971, -6.9648	12.5988, 15.9083, 20.736, 26.1851, 32.787

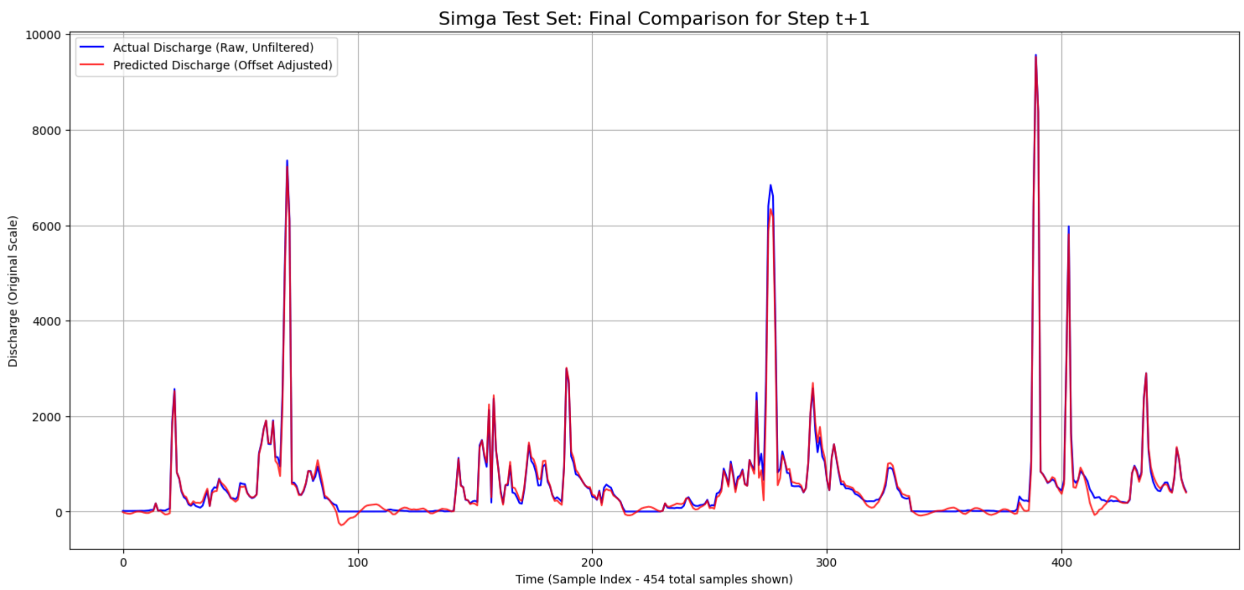
3. Simga

(a) Dataset Details & Plots

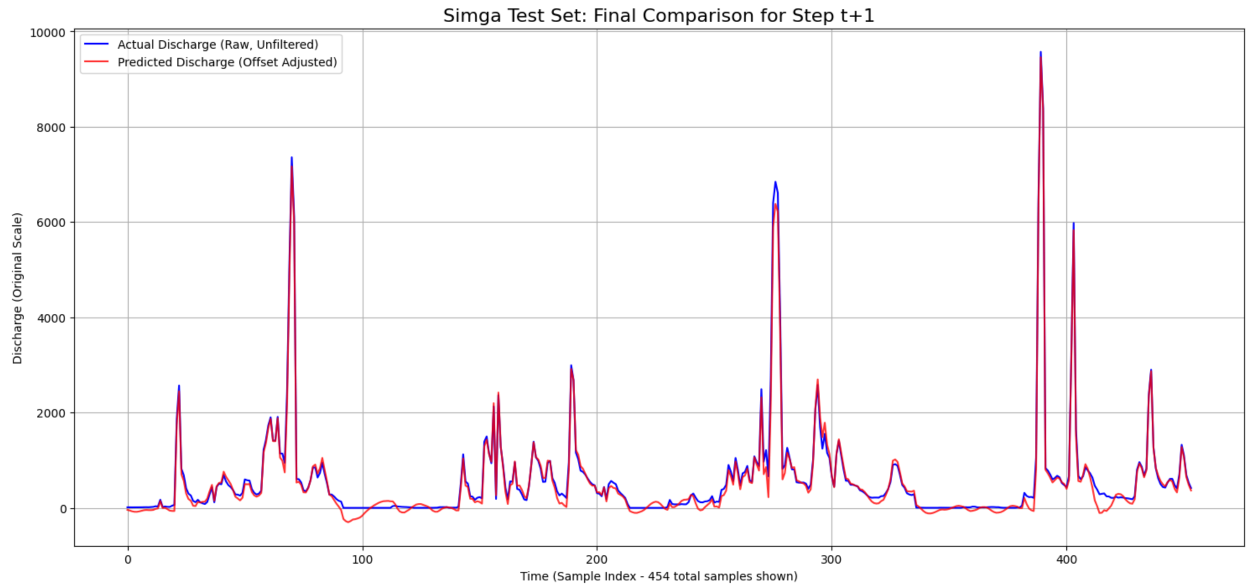
- **100% Data** (Train: [2000-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



- **80% Data** (Train: [2002-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



- **50% Data** (Train: [2005-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



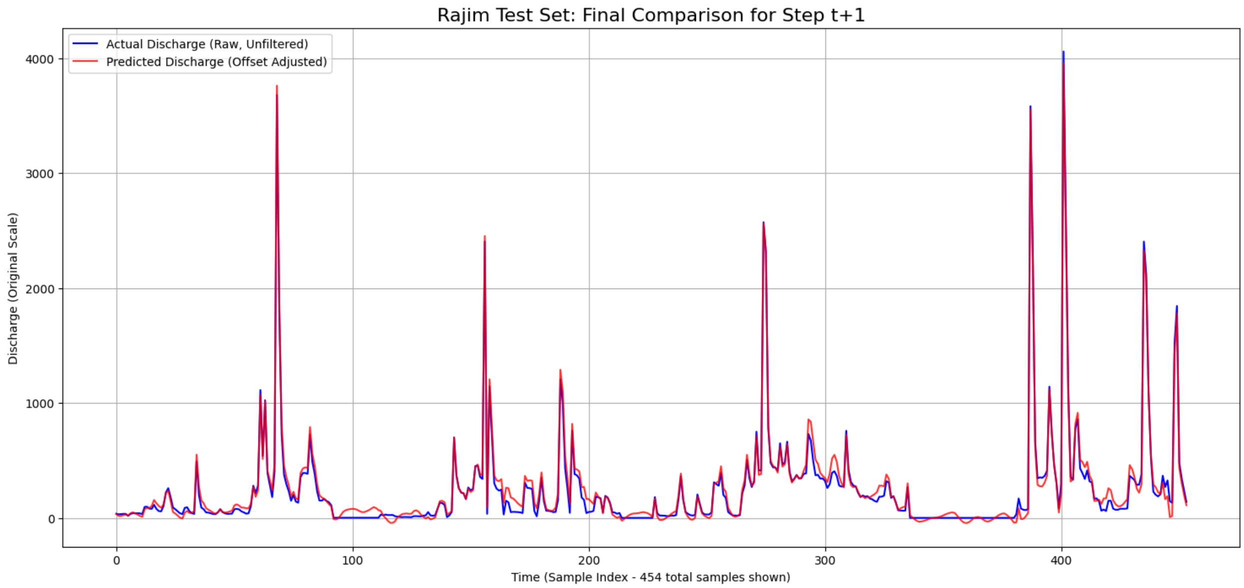
(b) Performance Metrics

Percentage	NSE Values	RSR Values	PBIAS (%)	TPE (%)
50%	[0.9919 0.9801 0.9591 0.9273 0.8843]	[0.0898 0.1409 0.2021 0.2695 0.3401]	[-2.3411 -2.2694 -0.9523 -1.2365 0.5535]	[11.2155 17.2471 24.0551 31.8188 40.8625]
80%	[0.9936 0.9819 0.9609 0.9303 0.8896]	[0.0803 0.1345 0.1978 0.2639 0.3323]	[-1.4183 -2.2446 -3.6909 -5.7752 -7.4983]	[10.2359 16.1525 22.3997 28.9535 36.1634]
100%	[0.9897 0.9736 0.9521 0.9162 0.8735]	[0.1015 0.1624 0.2188 0.2895 0.3557]	[-7.6407 -8.4787 -8.0628 -9.1254 -8.8749]	[13.6754 20.1432 25.7358 33.1292 40.4718]

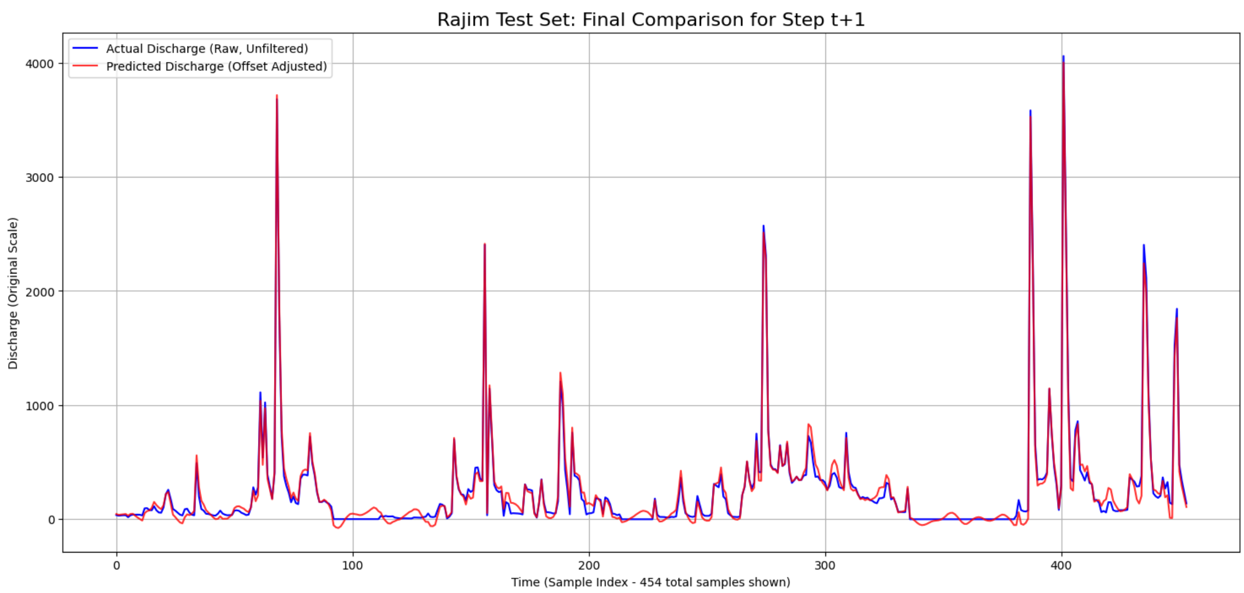
4. Rajim

(a) Dataset Details & Plots

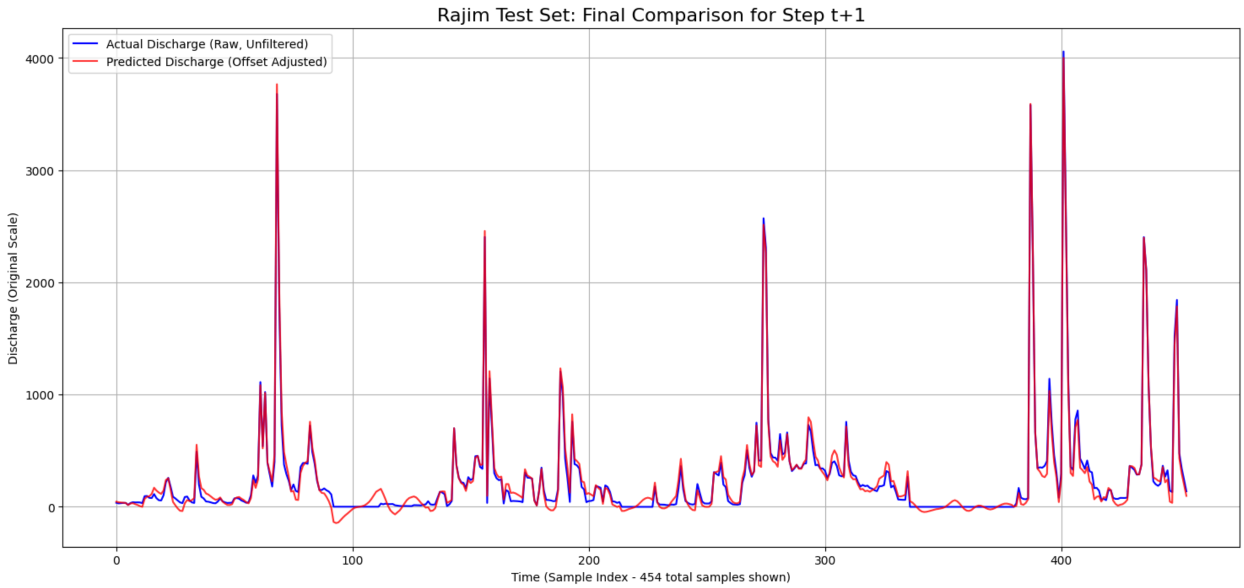
- **100% Data** (Train: [2000-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



- **80% Data** (Train: [2002-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



- **50% Data** (Train: [2005-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



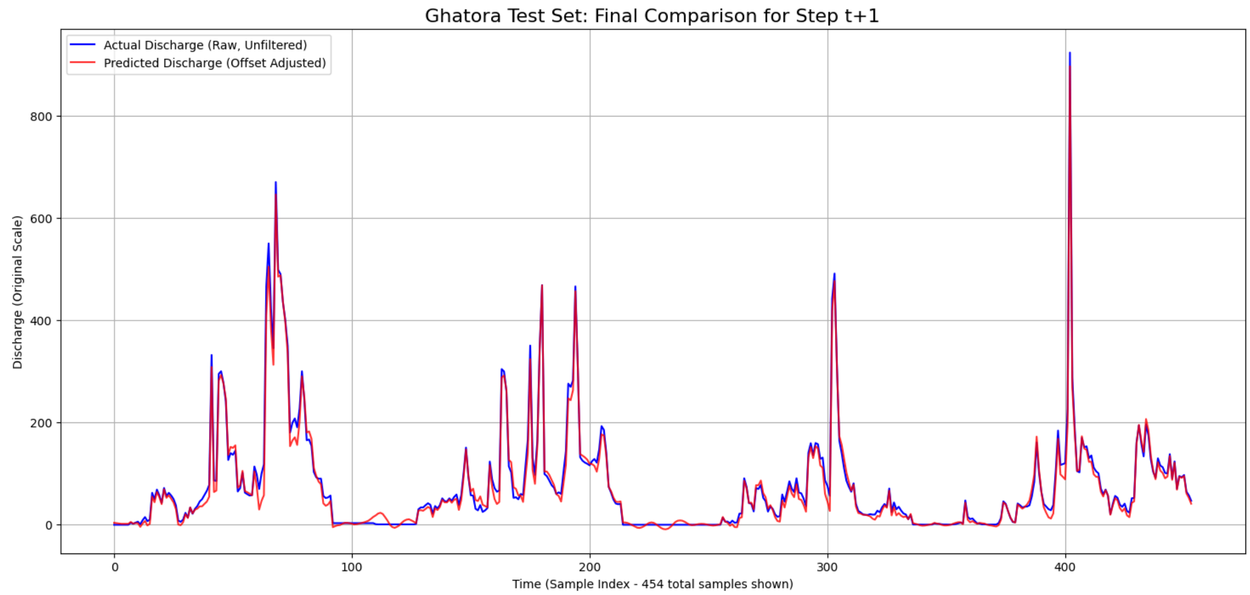
(b) Performance Metrics

Percentage	NSE Values	RSR Values	PBIAS (%)	TPE (%)
50%	[0.9884 0.9752 0.9504 0.9130 0.8633]	[0.1079 0.1576 0.2228 0.2949 0.3697]	[0.7890 6.1147 8.9366 14.1588 18.3437]	[15.9248 22.9493 32.6254 43.2297 54.0115]
80%	[0.9881 0.9762 0.9535 0.9168 0.8700]	[0.1092 0.1543 0.2157 0.2885 0.3606]	[6.5810 8.9223 11.6461 17.3474 21.6461]	[15.9538 22.1882 30.8207 41.5835 51.8656]
100%	[0.9888 0.9764 0.9535 0.9172 0.8711]	[0.1058 0.1537 0.2155 0.2878 0.3590]	[-0.1416 1.5246 5.2006 9.0877 13.5799]	[15.9882 22.8214 31.7123 41.9453 52.3063]

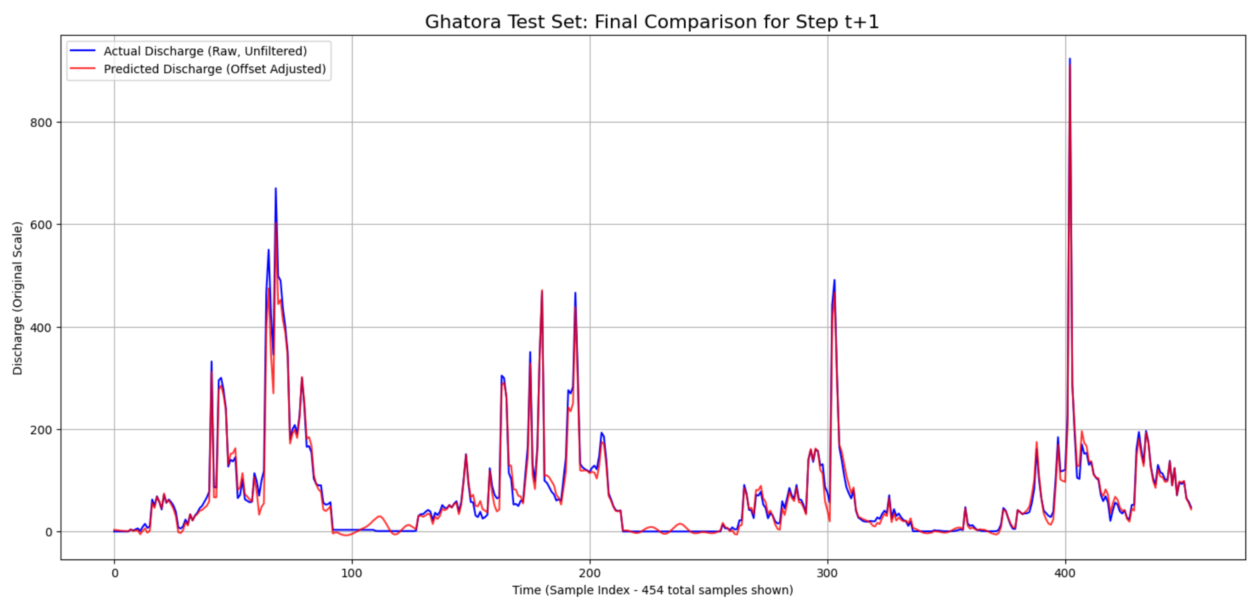
5. Ghatora

(a) Dataset Details & Plots

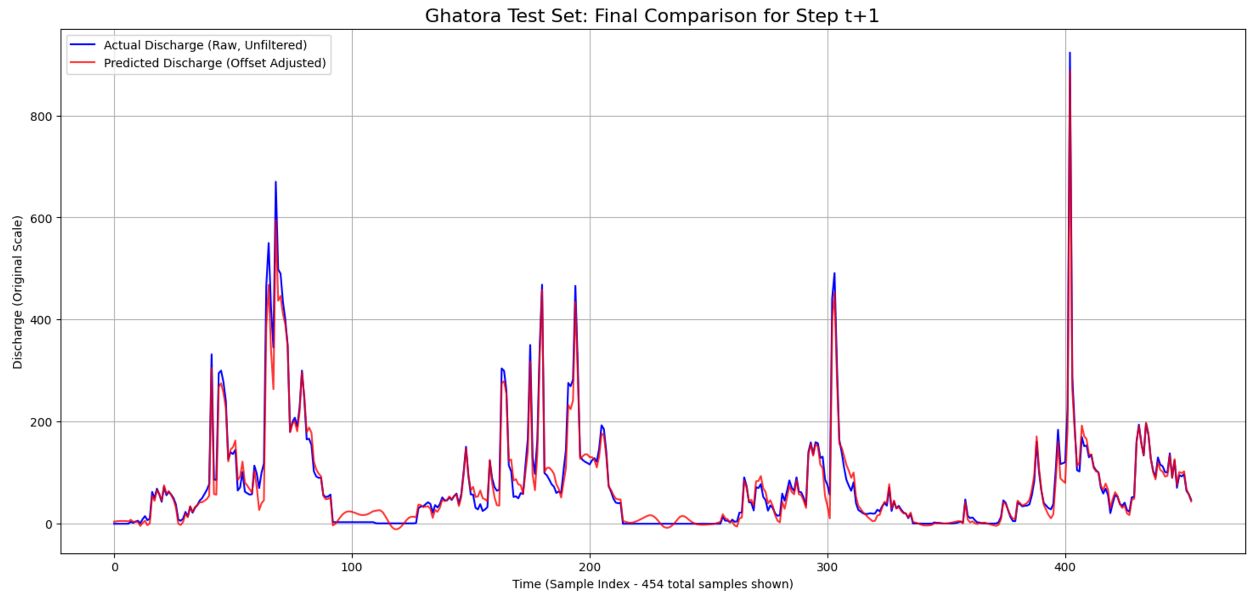
- **100% Data** (Train: [2000-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



- **80% Data** (Train: [2002-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



- **50% Data** (Train: [2005-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



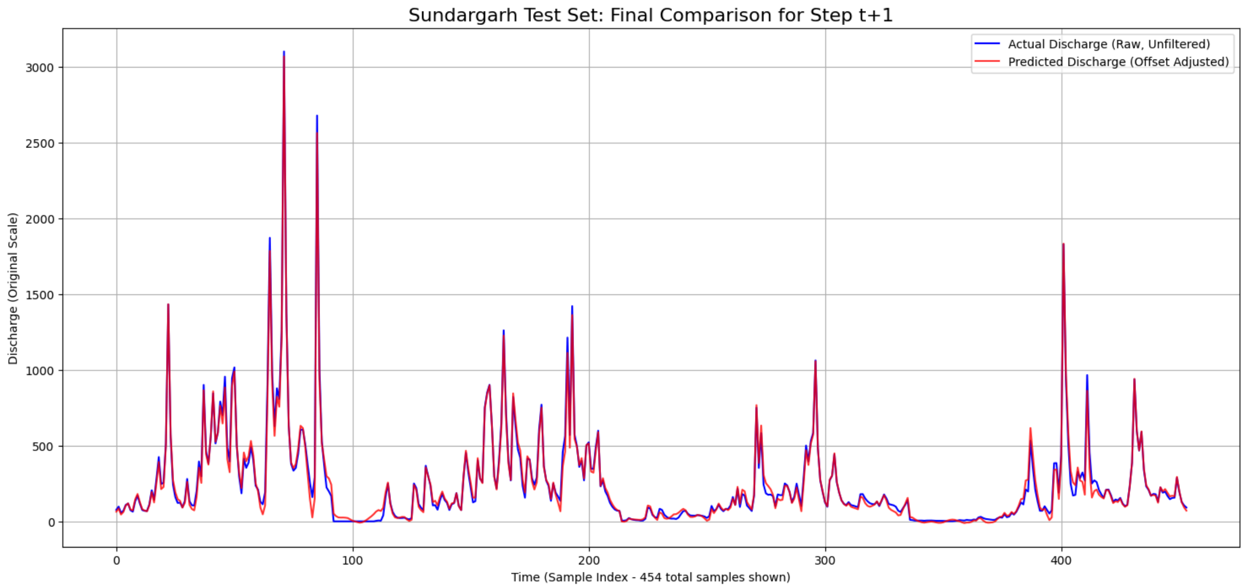
(b) Performance Metrics

Training Data %	NSE Values (t+1 to t+5)	RSR Values (t+1 to t+5)	PBIAS (%) (t+1 to t+5)	TPE (%) (t+1 to t+5)
100%	[0.9876 0.9686 0.9366 0.8974 0.8431]	[0.1111 0.1771 0.2518 0.3204 0.3961]	[-5.2629 -6.4444 -8.0281 -9.2135 -11.2116]	[10.7275 16.2422 21.9654 27.6858 33.6978]
80%	[0.9821 0.9638 0.9300 0.8905 0.8405]	[0.1340 0.1903 0.2645 0.3308 0.3993]	[-3.5486 -4.2369 -6.1835 -8.4911 -10.1819]	[12.2843 17.0515 23.0772 28.2154 33.8036]
50%	[0.9746 0.9519 0.9109 0.8643 0.8124]	[0.1594 0.2193 0.2985 0.3683 0.4332]	[-2.7701 -4.8514 -6.9663 -10.9531 -14.2436]	[14.7283 19.4066 25.7715 30.5392 35.1929]

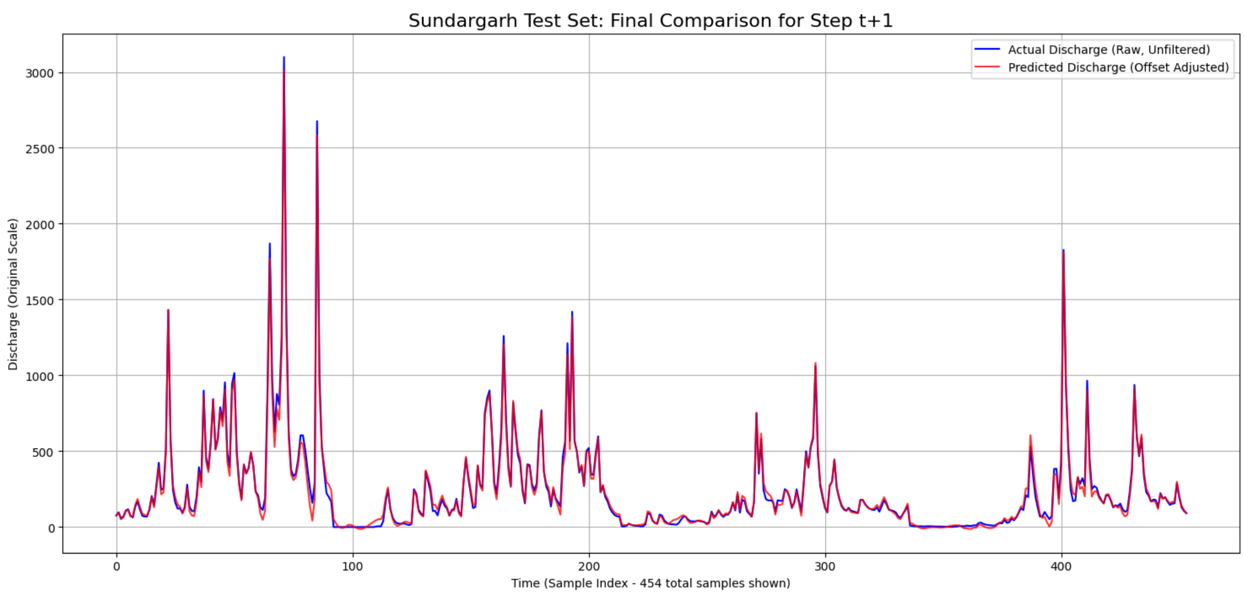
6. Sundargarh

(a) Dataset Details & Plots

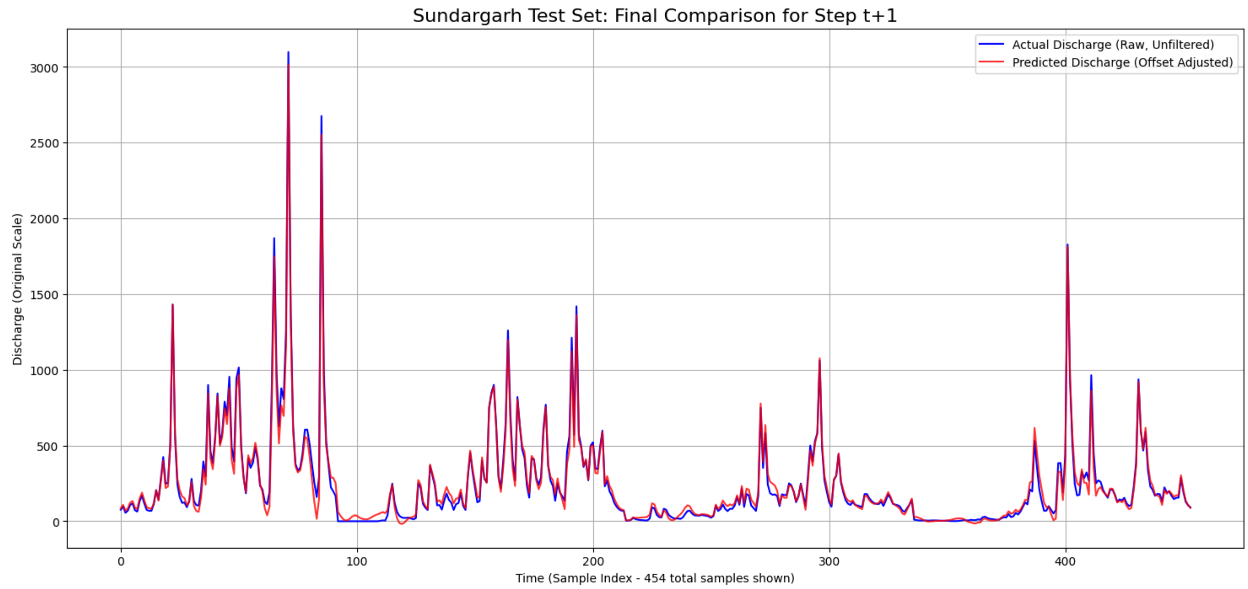
- **100% Data** (Train: [2000-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



- **80% Data** (Train: [2002-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



- **50% Data** (Train: [2005-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



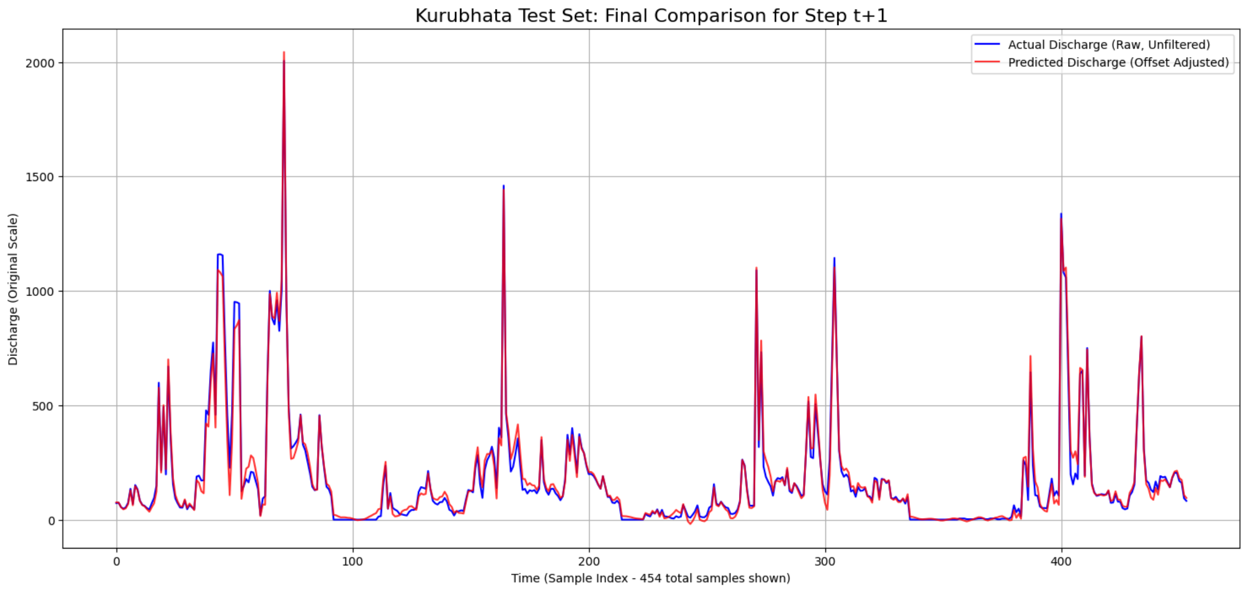
(b) Performance Metrics

Training Data %	NSE Values (t+1 to t+5)	RSR Values (t+1 to t+5)	PBIAS (%) (t+1 to t+5)	TPE (%) (t+1 to t+5)
100%	[0.9902 0.9761 0.9557 0.9230 0.8842]	[0.0988 0.1547 0.2105 0.2774 0.3404]	[-1.5029 -0.7051 0.3450 1.2732 2.7601]	[8.8521 14.0072 19.3585 25.5949 31.3205]
80%	[0.9916 0.9776 0.9548 0.9232 0.8822]	[0.0918 0.1496 0.2125 0.2772 0.3432]	[-2.0034 -2.2154 -1.2951 -2.0537 -2.0794]	[8.3322 13.5611 19.3089 25.2296 31.4858]
50%	[0.9865 0.9690 0.9431 0.9066 0.8656]	[0.1161 0.1762 0.2385 0.3056 0.3666]	[-0.5038 -2.0322 -2.6392 -4.9706 -6.5529]	[11.1401 16.2991 21.8378 27.2760 32.6907]

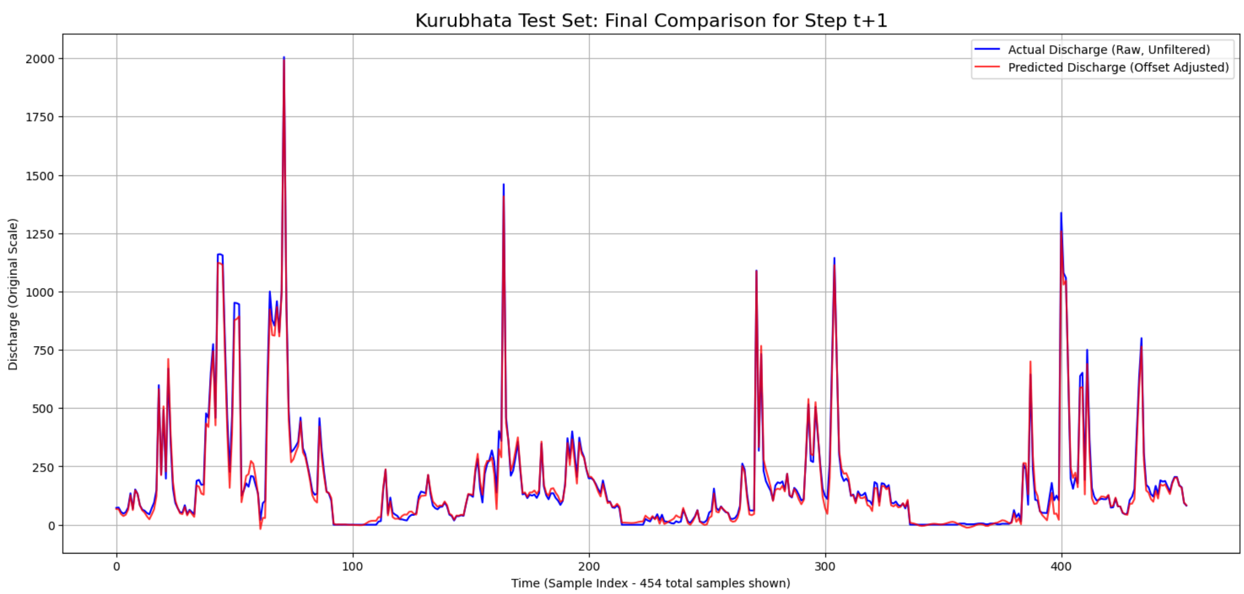
7. Kurubhata

(a) Dataset Details & Plots

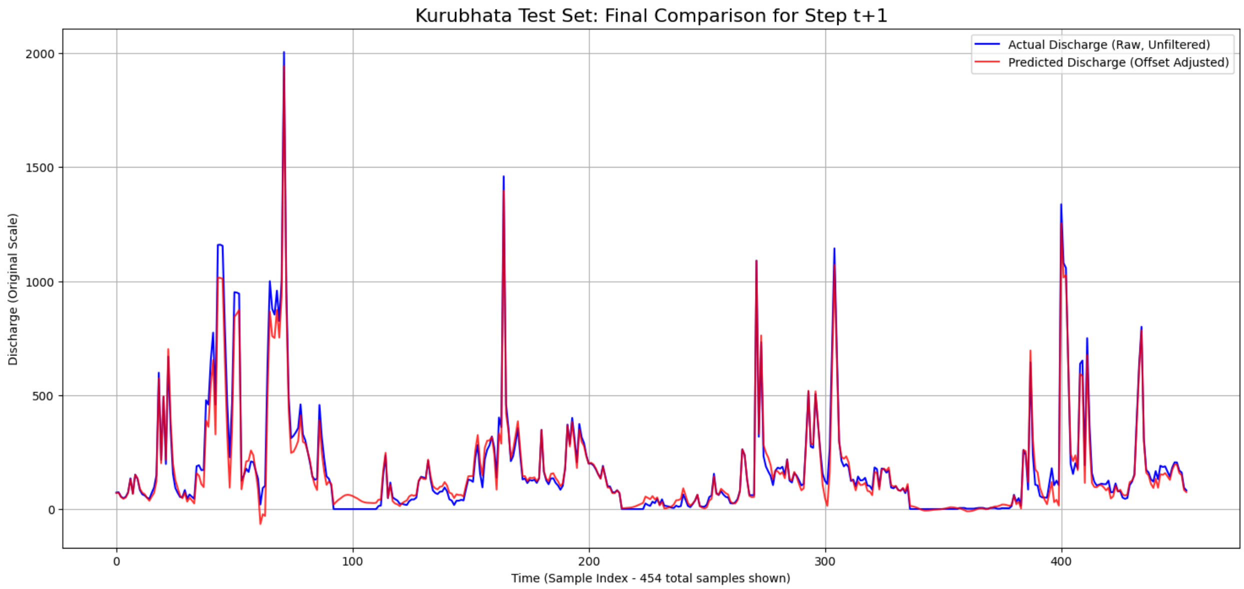
- **100% Data** (Train: [2000-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



- **80% Data** (Train: [2002-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



- **50% Data** (Train: [2005-06-01 00:00:00 to 2010-09-30], Test: [2011-06-01 00:00:00 to 2014-09-30 00:00:00])



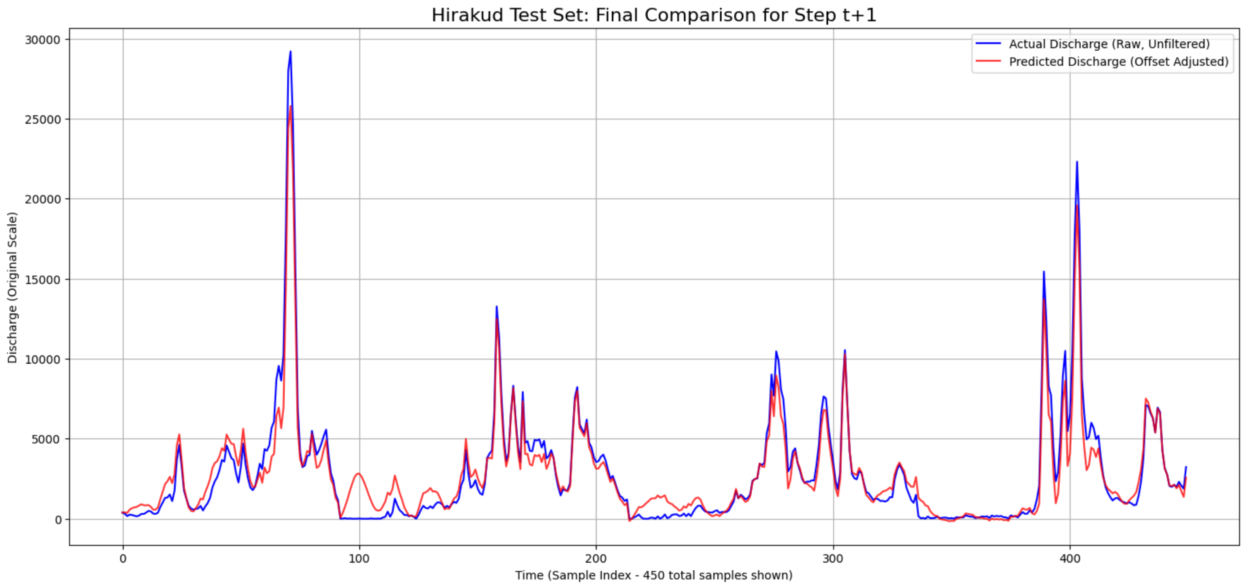
(b) Performance Metrics

Training Data %	NSE Values (t+1 to t+5)	RSR Values (t+1 to t+5)	PBIAS (%) (t+1 to t+5)	TPE (%) (t+1 to t+5)
100%	[0.9901 0.9745 0.9487 0.9153 0.8691]	[0.0994 0.1596 0.2265 0.2911 0.3618]	[-3.6103 -3.1558 -1.2200 -0.0669 1.0099]	[9.8176 15.0771 21.2779 27.4270 34.1240]
80%	[0.9868 0.9716 0.9459 0.9088 0.8617]	[0.1150 0.1686 0.2326 0.3020 0.3719]	[1.0213 1.0972 0.4419 -0.0252 0.3565]	[10.6403 16.0173 22.0030 27.8482 34.2533]
50%	[0.9761 0.9500 0.9125 0.8624 0.8002]	[0.1545 0.2236 0.2958 0.3710 0.4470]	[-4.0300 -5.6057 -7.6366 -10.1071 -13.2011]	[14.0864 19.8766 25.9660 31.8409 37.6178]

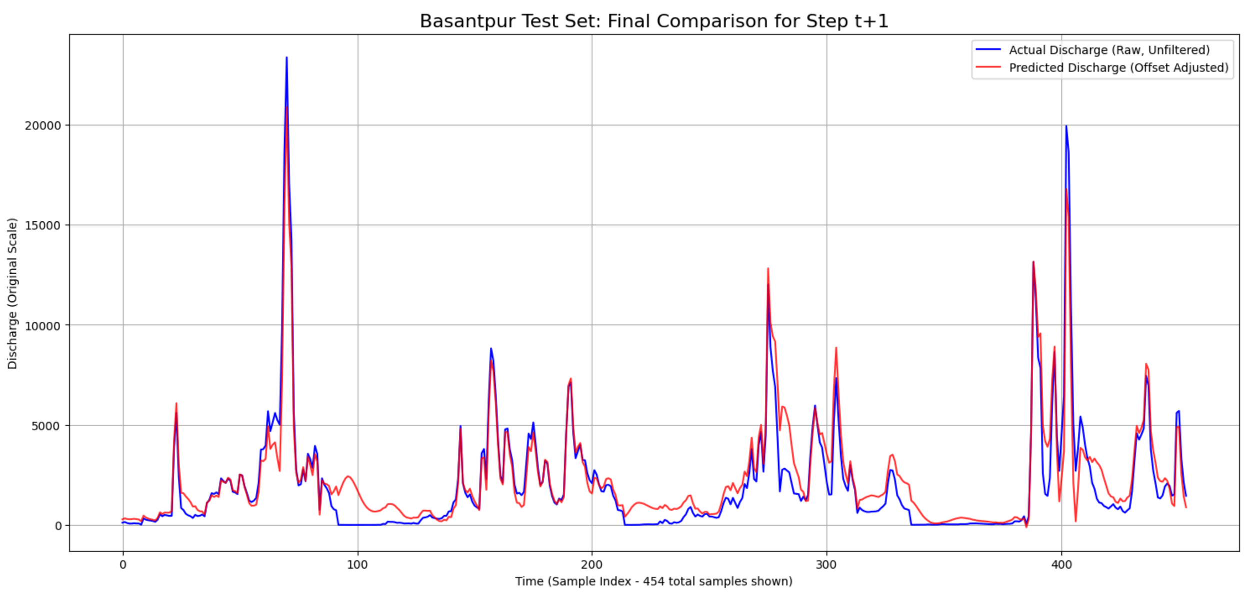
Additional Results with 20% Training Data

(Declining Results)

Station: ***Hirakud***



Station: *Basantpur*



Performance Metrics

Site	Step	NSE	RSR	PBIAS (%)	TPE (%)
Basantpur	t+1	0.9007	0.3152	14.3805	30.8062
	t+2	0.8711	0.3590	6.6045	34.6205
	t+3	0.8127	0.4328	4.2572	40.5732

	t+4	0.7513	0.4987	0.4886	48.5300
	t+5	0.6728	0.5720	0.2186	55.5251
Hirakud	t+1	0.9392	0.2465	-0.9962	22.3430
	t+2	0.9252	0.2734	-2.5418	24.1225
	t+3	0.8771	0.3505	-6.1698	30.5972
	t+4	0.8078	0.4384	-9.8458	37.1473
	t+5	0.7717	0.4778	-11.1664	41.4478