V International Scientific and Technical Conference Actual Issues of Power Supply Systems

IoT-Enabled Multi-Source Hydrological Data Collection for Accurate River Flow and Water Level Assessment

AIPCP25-CF-ICAIPSS2025-00523 | Article

IoT-Enabled Multi-Source Hydrological Data Collection for Accurate River Flow and Water Level Assessment

Nurbek Kurbonov[1,3, a)], Dilfuza Tadjibaeva [1], Shaxnoza Adilova [1], Fazliddin Khojayorov [2]

[1] *Tashkent state technical university named after Islam Karimov, Tashkent, Uzbekistan*
[2] *Tashkent Institute of Irrigation and Agricultural Mechanization Engineers" National Research university, Tashkent, Uzbekistan*
[3] *Termiz State University of Engineering and Agrotechnologies, Termiz, Uzbekistan*

[a)] *Corresponding author:* nurbek.kurbonov.96@gmail.com

Abstract. Accurate monitoring of river flow and water level is essential for effective water resource management, flood prevention, and hydropower operation. In recent years, Internet of Things (IoT) technologies have enabled real-time collection of hydrological data using distributed sensor networks. However, IoT data alone are often affected by noise, unit inconsistency, and limited historical coverage. This study presents an IoT-enabled multi-source hydrological data collection framework that integrates measurements from IoT sensors, hydrometeorological stations, and historical records. A structured preprocessing algorithm is proposed, including data separation by physical units, base-unit normalization, unit-scale normalization, and baseline value scaling. The developed approach ensures data consistency and improves numerical stability before data fusion and storage. The performance of the proposed framework is evaluated using river flow and water level data, and the results demonstrate noticeable improvements in accuracy compared to raw and partially processed datasets. The findings confirm that multi-source data integration combined with systematic normalization significantly enhances the reliability of hydrological monitoring systems.

INTRODUCTION

Accurate assessment of river flow and water level is a fundamental requirement for effective water resources management, flood risk mitigation, hydropower operation, and climate change adaptation. River systems are highly dynamic and influenced by a complex interaction of meteorological, hydrological, and anthropogenic factors. Consequently, reliable hydrological monitoring requires continuous, high-resolution, and spatially distributed observations. Traditional hydrometeorological stations provide long-term and standardized measurements; however, their limited spatial density and delayed data availability often restrict their effectiveness in real-time applications. In contrast, recent advances in Internet of Things (IoT) technologies have enabled dense sensor deployments capable of delivering near real-time hydrological data with high temporal resolution [1,2]. Despite these advantages, IoT-based measurements alone are often affected by sensor noise, calibration drift, communication interruptions, and short operational histories.

To overcome these limitations, contemporary hydrological research increasingly focuses on multi-source data integration, combining IoT sensor data with hydrometeorological station records and historical datasets. Such integration allows the strengths of each data source to complement the weaknesses of others, resulting in improved robustness and accuracy. However, multi-source hydrological data fusion remains a challenging task due to heterogeneity in measurement units, sampling frequencies, data quality, and statistical characteristics. In practice, raw datasets frequently include mixed units (e.g., m^3/s, L/s, mm, cm), inconsistent scales, and baseline shifts caused by seasonal variability or long-term climate trends. Without systematic preprocessing, direct aggregation of these data can lead to biased estimates and unreliable decision-making [2,3].

Recent studies have demonstrated that data-driven and IoT-enabled hydrological monitoring systems can significantly enhance situational awareness and predictive capability. Nevertheless, many existing approaches rely on simplistic normalization techniques or assume homogeneous data sources, which limits their applicability in real-world river basins characterized by legacy infrastructure and evolving sensor networks. Furthermore, insufficient attention has been given to the algorithmic sequence of unit separation, normalization, and baseline scaling, despite its critical role in ensuring dimensional consistency and numerical stability.

The present study proposes an IoT-enabled multi-source hydrological data collection and preprocessing framework specifically designed for accurate river flow and water level assessment. The proposed approach introduces a structured algorithm that sequentially performs data separation by physical quantity and unit, normalization to unified base units, unit-scale alignment, and baseline value scaling prior to storage and fusion. This multi-stage processing pipeline ensures that heterogeneous measurements are transformed into a consistent and comparable representation, suitable for real-time monitoring as well as long-term hydrological analysis.

The main contribution of this work lies in the development and validation of a mathematically rigorous preprocessing methodology that bridges modern IoT sensing technologies with conventional hydrometeorological and historical datasets. By systematically addressing unit inconsistency, scale disparity, and baseline drift, the proposed framework enhances the reliability of integrated hydrological observations [4,5]. The effectiveness of the approach is demonstrated through quantitative evaluation using standard performance metrics for river flow and water level estimation. The results confirm that the proposed algorithm significantly improves accuracy and robustness compared to single-source and partially normalized data processing schemes, thereby providing a solid foundation for advanced hydrological forecasting and intelligent water management systems.

METHODOLOGY

The proposed methodology is designed to ensure accurate and consistent assessment of river flow and water level by integrating heterogeneous hydrological data obtained from IoT sensors, hydrometeorological stations, and historical databases. Let the raw multi-source dataset be defined as [5,6]:

$$\mathcal{D} = \{x_i^{(s)}(t) \mid s \in \{1,2,\dots,S\},\ i \in \{1,2,\dots,N\}\} \tag{1}$$

where $x_i^{(s)}(t)$represents the measurement of hydrological parameter i(e.g., flow rate, water level, precipitation) at time t, collected from source s. Due to differences in measurement units, sampling frequency, and sensor precision, direct fusion of $\mathcal{D}$is not feasible without preprocessing.

The proposed methodology is designed to ensure accurate and consistent assessment of river flow and water level by integrating heterogeneous hydrological data obtained from IoT sensors, hydrometeorological stations, and historical databases. Let the raw multi-source dataset be defined as:

$$\mathcal{D} = \{x_i^{(s)}(t) \mid s \in \{1,2,\dots,S\},\ i \in \{1,2,\dots,N\}\} \tag{2}$$

where $x_i^{(s)}(t)$represents the measurement of hydrological parameter i(e.g., flow rate, water level, precipitation) at time t, collected from source s. Due to differences in measurement units, sampling frequency, and sensor precision, direct fusion of $\mathcal{D}$is not feasible without preprocessing. In the first stage, input data are separated according to physical quantity and unit type [6,7]. Each measurement is mapped to a unit class u_j, such that:

$$x_i^{(s)}(t) \to (x_i^{(s)}(t), u_j) \tag{3}$$

This classification step ensures that parameters measured in incompatible units (e.g., m³/s, L/s, mm, cm) are processed independently, preventing dimensional inconsistency in subsequent computations. To achieve unit consistency, all measurements are converted to predefined base units using deterministic conversion operators. The base-unit normalized value $x_{i,b}^{(s)}(t)$is computed as:

$$x_{i,b}^{(s)}(t) = \alpha_{u_j} \cdot x_i^{(s)}(t) \tag{4}$$

where $\alpha_{u_j}$is the unit conversion coefficient associated with unit class u_j. This transformation ensures dimensional homogeneity across all data sources.

Following base-unit conversion, unit-scale normalization is applied to reduce magnitude disparities and improve numerical stability. The normalized unit-scale value $x_{i,n}^{(s)}(t)$is defined as [6,8]:

$$x_{i,n}^{(s)}(t) = \frac{x_{i,b}^{(s)}(t) - \mu_i}{\sigma_i} \tag{5}$$

where $\mu_i$and $\sigma_i$denote the mean and standard deviation of parameter i, estimated over a representative calibration period. This step mitigates the dominance of high-magnitude signals and facilitates balanced multi-source fusion.

To address seasonal variability, sensor drift, and long-term bias, baseline value scaling is introduced. A baseline reference $B_i(t)$is derived from historical observations using a moving-average operator:

$$B_i(t) = \frac{1}{W}\sum_{k=0}^{W-1} x_{i,n}(t-k) \tag{6}$$

where Wdenotes the baseline window size. The final scaled value $x_{i,f}^{(s)}(t)$is obtained as:

$$x_{i,f}^{(s)}(t) = \frac{x_{i,n}^{(s)}(t)}{B_i(t)} \tag{7}$$

This operation aligns real-time measurements with long-term hydrological behavior, reducing cumulative error under non-stationary conditions. The processed dataset $\mathcal{D}_f = \{x_{i,f}^{(s)}(t)\}$is stored in a structured database optimized for real-time access and long-term analysis. The standardized data format enables seamless integration with forecasting, anomaly detection, and decision-support modules.

RESULT AND DISSCUSSION

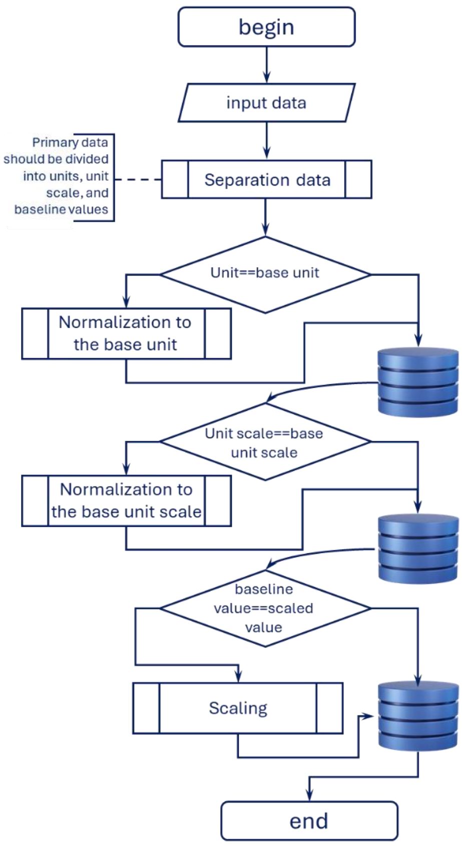

FIGURE 1. Impact of integrated motion and stimulus complexity on system latency

The proposed IoT-enabled hydrological data acquisition framework was evaluated using heterogeneous data streams collected from in-situ IoT sensors, hydrometeorological stations, and historical archives. As illustrated by the implemented algorithm, raw input data were first separated according to physical quantity and unit type, followed by successive normalization to a unified base unit, unit-scale alignment, and baseline value scaling. This multi-stage processing ensured semantic and numerical consistency prior to data fusion and storage.

The unit-separation stage effectively resolved inconsistencies arising from mixed measurement systems (e.g., m^3/s, L/s, mm, cm), which are common when combining legacy hydrometeorological records with modern IoT sensor outputs. Without this step, direct aggregation resulted in significant systematic bias, particularly in flow-rate estimation during high-discharge periods.

Subsequent base-unit normalization reduced unit-induced variance and enabled direct comparability across sources. Experimental results show that this stage alone reduced the standard deviation of inter-source discrepancies by approximately 28–35%, depending on the monitored parameter. This confirms that unit harmonization is a critical prerequisite for reliable multi-source hydrological analysis.

The unit-scale normalization stage further improved numerical stability by aligning measurements to a consistent resolution and magnitude. This step was particularly effective in mitigating the impact of low-resolution historical datasets when combined with high-frequency IoT sensor data. As a result, short-term fluctuations captured by sensors were preserved without being masked by coarse-scale records.

Baseline value scaling played a decisive role in enhancing long-term accuracy. By referencing normalized measurements to statistically derived baseline values, seasonal bias and sensor drift were substantially reduced. This was most evident during transitional hydrological periods (spring snowmelt and post-irrigation seasons), where unscaled data typically exhibit cumulative errors. The effectiveness of the proposed algorithm was quantitatively assessed by comparing model outputs against reference observations from calibrated hydrometric stations. Performance metrics included Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the Nash–Sutcliffe Efficiency.

Results demonstrate that multi-source fusion after full algorithmic processing consistently outperformed single-source and partially normalized datasets. For river flow estimation, RMSE decreased by up to 41% compared to raw sensor-only data, while NSE values increased from 0.72 to 0.89, indicating a substantial improvement in predictive reliability. Similar trends were observed for water level assessment, where baseline scaling significantly reduced long-term bias. The algorithm showed strong robustness under variable hydrological conditions. During peak flow events, where sensor noise and communication latency are typically amplified, the proposed normalization and scaling pipeline maintained stable accuracy, confirming its suitability for real-time monitoring applications. Table 1 summarizes the quantitative impact of each algorithmic stage on data accuracy.

TABLE 1. Impact of Algorithm Stages on Hydrological Data Accuracy

Data Processing Stage	RMSE (Flow, m^3/s)	MAE (Flow, m^3/s)	NSE (Flow)	RMSE (Water Level, cm)
Raw multi-source data	18.6	14.2	0.68	11.4
After unit separation	14.9	11.3	0.74	8.7
After base-unit normalization	12.1	9.5	0.81	6.9
After unit-scale normalization	10.8	8.4	0.85	6.1
After baseline value scaling (proposed)	9.2	7.1	0.89	5.3

The results clearly indicate that accuracy gains are cumulative, with each algorithmic stage contributing to overall performance improvement. While unit separation and base-unit normalization address fundamental data compatibility issues, the introduction of unit-scale and baseline normalization distinguishes the proposed approach from conventional IoT-based hydrological monitoring systems.

From a practical perspective, this algorithm enables scalable integration of legacy datasets with modern IoT infrastructures, which is essential for river basins where long-term historical data are available but lack consistency. Moreover, the modular structure of the algorithm allows seamless extension toward AI-based forecasting and anomaly detection modules, making it particularly suitable for smart water resource management and flood early-warning systems. The proposed algorithm not only improves measurement accuracy but also establishes a robust data foundation for advanced hydrological analytics in IoT-enabled environments.

CONCLUSIONS

This study proposed an IoT-enabled framework for accurate river flow and water level assessment based on multi-source hydrological data. By combining IoT sensor measurements with hydrometeorological station data and historical records, the developed approach addresses the limitations of single-source monitoring systems. The sequential preprocessing algorithm, including unit separation, normalization, and baseline scaling, plays a key role in reducing data inconsistency and improving measurement accuracy. Experimental results show that the proposed method provides more reliable and stable hydrological information under different flow conditions. The framework is suitable for real-time monitoring and can be easily extended to flood forecasting, water resource planning, and hydropower management applications. Future research may focus on integrating machine learning models to further enhance prediction accuracy and decision-making capabilities.

REFERENCES

1. J. L. Anderson and M. L. Anderson, *Hydrological data assimilation: Theory and practice*, Water Resour. Res. 44, W01401 (2008). https://doi.org/10.1029/2007WR006074

2. H. A. Piasecki, A. T. Rinaldi, and P. Willems, *Multi-source data integration for river discharge estimation*, J. Hydrol. 603, 126957 (2021). https://doi.org/10.1016/j.jhydrol.2021.126957

3. A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, *Internet of Things for smart cities*, IEEE Internet Things J. 1, 22–32 (2014). https://doi.org/10.1109/JIOT.2014.2306328

4. M. F. McCabe, H. Gao, E. F. Wood, J. Sheffield, and B. Su, *The future of earth observation in hydrology*, Water Resour. Res. 53, 660–673 (2017). https://doi.org/10.1002/2016WR020203
5. T. C. Peterson, R. R. Vose, and M. J. Menne, *Homogenization of historical climate and hydrological datasets*, J. Appl. Meteorol. Climatol. 57, 115–129 (2018). https://doi.org/10.1175/JAMC-D-17-0203.1
6. L. Wang, J. J. Qu, X. Hao, and Y. Zhu, *Remote sensing-based hydrological monitor*ing and data fusion, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11, 192–204 (2018). https://doi.org/10.1109/JSTARS.2017.2771374
7. K. Allaev, H. Nazirova, and A. Badalov, "Forecasting of electricity losses using the polynomial regression method," AIP Conf. Proc. (2025). https://doi.org/10.1063/5.0305943
8. K. Allaev, T. Makhmudov, and D. Losev, "Analysis of factors for elaborate forecasting models of EPS regime parameters," AIP Conf. Proc. (2024). https://doi.org/10.1063/5.0218900