ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ КАЗАХСТАН
MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE REPUBLIC OF KAZAKHSTAN

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

2025 (23) 3

шілде- қыркүйек

ISSN 2708–2032 (print)
ISSN 2708–2040 (online)

Халықаралық ақпараттық және коммуникациялық технологиялар журналы
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Меншік иесі: АҚ «Халықаралық ақпараттық технологиялар университеті» (Алматы қ.).
Қазақстан Республикасы Ақпарат және қоғамдық даму министрлігіне мерзімді баспасөз басылымын есепке қою туралы куәлік № KZ82VPY00020475, 20.02.2020 ж. берілген
Тақырып бағыты: ақпараттық технологиялар, ақпараттық қауіпсіздік және коммуникациялық технологиялар, әлеуметтік-экономикалық жүйелерді дамытудағы цифрлық технология.
Мерзімділігі: жылына 4 рет.
Тираж: 100 дана.
Редакция мекенжайы: 050040 Алматы қ., Манас к., 34/1, каб. 709, тел: +7 (727) 244-51-09.
E-mail: ijict@iitu.edu.kz
Журнал сайты: https://journal.iitu.edu.kz

Международный журнал информационных и коммуникационных технологий
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Собственник: АО «Международный университет информационных технологий» (г. Алматы).
Свидетельство о постановке на учет периодического печатного издания в Министерство информации и общественного развития Республики Казахстан **№ KZ82VPY00020475,** выданное от **20.02.2020 г.**
Тематическая направленность: информационные технологии, информационная безопасность и коммуникационные технологии, цифровые технологии в развитии социо-экономических систем**.**
Периодичность: 4 раза в год.
Тираж: 100 экземпляров.
Адрес редакции: 050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09.
E-mail: ijict@iitu.edu.kz Сайт журнала: https://journal.iitu.edu.kz

«International Journal of Information and Communication Technologies»
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Owner: International Information Technology University JSC (Almaty).
The certificate of registration of a periodical printed publication in the Ministry of Information and Social Development of the Republic of Kazakhstan, Information Committee No. KZ82VPY00020475, issued on 20.02.2020.
Thematic focus: information technology, digital technologies in the development of socio-economic systems, information security and communication technologies
Periodicity: 4 times a year.
Circulation: 100 copies.
Editorial address: 050040. Manas st. 34/1, Almaty. +7 (727) 244-51-09. E-mail: ijict@iitu.edu.kz
Journal website: https://journal.iitu.edu.kz

МАЗМҰНЫ

СОДЕРЖАНИЕ

CONTENTS

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 6. Is. 3. Number 23 (2025). Pp. 24–44
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2025.23.3.002
UDC 004.8

IMPROVING ROBUSTNESS IN AI FLOOD FORECASTING VIA BLOCKCHAIN-INSPIRED CONSENSUS MODELS

A. Alzhanov[1*], K. Rakhymbek[2], A. Nugumanova[1]
[1]Astana IT University, Astana, Kazakhstan;
[2]Sarsen Amanzholov East Kazakhstan University, Oskemen, Kazakhstan.
E-mail: almas.alzhanov01@gmail.com

Alzhanov Almas — PhD student, junior researcher, Science and Innovation Center "Big Data and Blockchain Technologies", Astana IT University, Astana, Kazakhstan
E-mail: almas.alzhanov01@gmail.com https://orcid.org/0009-0007-8083-2366;
Rakhymbek Kamilla — Master's student, junior researcher, Laboratory of Digital Technologies and Modeling, Sarsen Amanzholov East Kazakhstan University, Oskemen, Kazakhstan
https://orcid.org/0009-0008-7404-8433;
Nugumanova Aliya — PhD, Director of Science and Innovation Center "Big Data and Blockchain Technologies", Astana IT University, Astana, Kazakhstan
https://orcid.org/0000-0001-5522-4421.

Abstract. Accurate streamflow forecasting is critical for effective water resource management and flood mitigation. While ensemble forecasting improves robustness, its potential is often limited by static aggregation techniques that fail to leverage the dynamic reliability of individual models, often dampening the signal from the most accurate forecast. This study addresses this gap by presenting and evaluating a model-agnostic ensemble framework inspired by decentralized consensus mechanisms in blockchain technology, designed to enhance forecast accuracy and robustness. The framework integrates daily predictions from four distinct Long Short Term Memory models using three dynamic aggregation strategies: Quorum-based Median Agreement, Skill-Weighted Voting, and Adaptive Leader Selection. For comparison, we also evaluate established adaptive ensemble methods, namely Online Super Learner and Dynamic Model Averaging. The blockchain-inspired strategies treat each model as an independent node, reaching a collective agreement based on dynamic performance metrics. To further improve operational reliability, an online, no-leakage debiasing module was applied as a post-processing step to correct for systematic forecast errors. Experimental results show that the consensus strategies

outperform both the individual models, the traditional ensemble average, and the additional adaptive baselines. After debiasing, the Skill-Weighted Voting approach achieved the highest overall accuracy with a Kling-Gupta Efficiency of 0.965 and a Nash-Sutcliffe Efficiency of 0.933, while the Adaptive Leader Selection strategy proved most robust, attaining the lowest 90th percentile absolute error, thus reducing the magnitude of large forecast errors. These findings demonstrate that combining a blockchain-inspired consensus approach with real-time error correction provides a practical and effective pathway for developing more resilient forecasting models.

Key words: streamflow forecasting, blockchain, consensus mechanisms, LSTM, bias correction, flood forecasting, model robustness

For citation: A. Alzhanov, K. Rakhymbek, A. Nugumanova. Improving robustness in ai flood forecasting via blockchain-inspired consensus models // International journal of information and communication technologies. 2025. Vol. 6. No. 23. Pp. 24–44. (In Eng.). https://doi.org/10.54309/IJICT.2025.23.3.002.

Conflict of interest: The authors declare that there is no conflict of interest.

БЛОКЧЕЙННЕН ШАБЫТТАНҒАН КОНСЕНСУС МОДЕЛЬДЕРІ АРҚЫЛЫ ЖАСАНДЫ ИНТЕЛЛЕКТ НЕГІЗІНДЕГІ СУ ТАСҚЫНЫ БОЛЖАМЫНЫҢ ТҰРАҚТЫЛЫҒЫН АРТТЫРУ

А.М. Альжанов[1]*, К.Қ. Рахымбек[2], А.Б. Нугуманова[11]

Astana IT University, Астана, Қазақстан;

[2]Сәрсен Аманжолов атындағы Шығыс Қазақстан университеті, Өскемен, Қазақстан.

E-mail: almas.alzhanov01@gmail.com

Альжанов Алмас — докторант, «Big Data and Blockchain Technologies» ғылыми-инновациялық орталығының кіші ғылыми қызметкері, Astana IT University https://orcid.org/0009-0007-8083-2366;

Рахымбек Камилла — магистрант, Цифрлық Технологиялар және Модельдеу Зертханасының кіші ғылыми қызметкері, Сәрсен Аманжолов атындағы Шығыс Қазақстан университеті https://orcid.org/0009-0008-7404-8433;

Нугуманова Алия — PhD, «Big Data and Blockchain Technologies» ғылыми-инновациялық орталығының директоры, Astana IT University https://orcid.org/0000-0001-5522-4421.

Аннотация. Өзен ағынын дәл болжау су ресурстарын тиімді басқару және су тасқынын азайту үшін өте маңызды. Ансамбльді болжау сенімділікті арттырғанымен, оның әлеуеті көбінесе жеке модельдердің динамикалық сенімділігін пайдалана алмайтын статикалық біріктіру әдістерімен шектеледі,

бұл көбінесе сигналды ең дәл болжамнан әлсіретеді. Бұл зерттеу болжамдардың дәлдігі мен сенімділігін арттыруға арналған блокчейн технологиясындағы орталықтандырылмаған консенсус механизмдерінен шабыттанған модельдік-агностикалық ансамбль құрылымын ұсыну және бағалау арқылы осы олқылықты жояды. Бұл құрылым үш динамикалық біріктіру стратегиясын қолдана отырып, төрт түрлі Long Short Term Memory модельдерінің күнделікті болжамдарын біріктіреді: Quorum-based Median Agreement, Skill-Weighted Voting, және Adaptive Leader Selection. Салыстыру үшін, біз сондай-ақ ансамбльдің қалыптасқан бейімделу әдістерін, атап айтқанда Online Super Learner және Dynamic Model Averaging бағалаймыз. Блокчейннен шабыттандырылған стратегиялар әрбір модельді динамикалық өнімділік көрсеткіштеріне негізделген ұжымдық келісімге қол жеткізе отырып, тәуелсіз түйін ретінде қарастырады. Операциялық сенімділікті одан әрі жақсарту үшін жүйелі болжау қателерін түзету үшін өңдеуден кейінгі қадам ретінде online no-leakage debiasing модулі қолданылды. Эксперименттік нәтижелер консенсус стратегиялары жеке модельдерден де, ансамбльдің дәстүрлі статикалық орталау көрсеткішінен де, қосымша бэйслайндер де асып түсетінін көрсетеді. Debiasing модулінен шығарылғаннан кейін Skill-Weighted Voting тәсілі ең жоғары жалпы дәлдікке қол жеткізді, Kling-Gupta Efficiency 0.965 және Nash-Sutcliffe Efficiency 0.933 Болды, ал Adaptive Leader Selection стратегиясы ең сенімді болып шықты, абсолютті 90-шы процентильдегі ең төменгі қателікке қол жеткізді, осылайша болжамдардағы үлкен қателіктердің ауқымын азайтты. Бұл нәтижелер блокчейннен шабыттандырылған консенсус тәсілін нақты уақыттағы қателерді түзетумен біріктіру тұрақты болжау модельдерін әзірлеудің практикалық және тиімді жолын қамтамасыз ететінін көрсетеді.

Түйін сөздер: өзен ағынын болжау, блокчейн, консенсус механизмдері, LSTM, бұрмалануды түзету, су тасқынын болжау, модельдің беріктігі

Дәйексөз үшін: А.М. Альжанов, К.Қ. Рахымбек, А.Б. Нугуманова. Блокчейннен шабыттанған консенсус модельдері арқылы жасанды интеллект негізіндегі су тасқыны болжамының тұрақтылығын арттыру//Халықаралық ақпараттық және коммуникациялық технологиялар журналы. 2025. Том. 6. № 23. 24–44 бет. (Ағыл). https://doi.org/10.54309/IJICT.2025.23.3.002.

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

ПОВЫШЕНИЕ УСТОЙЧИВОСТИ ПРОГНОЗИРОВАНИЯ НАВОДНЕНИЙ НА ОСНОВЕ ИИ С ИСПОЛЬЗОВАНИЕМ КОНСЕНСУСНЫХ МОДЕЛЕЙ, ВДОХНОВЛЁННЫХ БЛОКЧЕЙНОМ

А.М. Альжанов[1]*, К.К. Рахымбек[2], А.Б. Нугуманова[11]

Астана ИТ университет, Астана, Казахстан;

[2]Восточно-Казахстанский университет имени С. Аманжолова, Усть-

Каменогорск, Казахстан.
E-mail: almas.alzhanov01@gmail.com

Альжанов Алмас — докторант, младший научный сотрудник научно-инновационного центра «Больших данных и технологии блокчейн», Астана ИТ университет, Астана, Казахстан
E-mail: almas.alzhanov01@gmail.com, https://orcid.org/0009-0007-8083-2366;
Рахымбек Камилла — магистрант, младший научный сотрудник лаборатории «Цифровых технологий и моделирования», Восточно-Казахстанский университет имени С. Аманжолова, Усть-Каменогорск, Казахстан
https://orcid.org/0009-0008-7404-8433;
Нугуманова Алия — PhD, директор научно-инновационного центра «Больших данных и технологии блокчейн»», Астана ИТ университет, Астана, Казахстан
https://orcid.org/0000-0001-5522-4421.

Аннотация. Точное прогнозирование стока (расхода воды) имеет решающее значение для эффективного управления водными ресурсами и снижения риска наводнений. Хотя ансамблевое прогнозирование повышает робастность, его потенциал нередко ограничивается статическими методами агрегирования, которые не учитывают динамическую надежность отдельных моделей и часто «приглушают» сигнал от наиболее точного прогноза. В данной работе этот разрыв закрывается за счет представления и оценки модель-независимого ансамблевого фреймворка, вдохновленного децентрализованными механизмами консенсуса в технологии блокчейн и предназначенного для повышения точности и робастности прогнозов. Фреймворк интегрирует ежедневные предсказания четырех различных моделей Long Short Term Memory, используя три динамические стратегии агрегирования: Quorum-based Median Agreement, Skill-Weighted Voting и Adaptive Leader Selection. Для сравнения также оцениваются известные адаптивные ансамблевые методы — Online Super Learner и Dynamic Model Averaging. Стратегии, вдохновленные блокчейном, рассматривают каждую модель как независимый узел, приходящий к коллективному соглашению на основе динамических метрик качества. Для дальнейшего повышения операционной надежности в качестве шага постобработки применен модуль для коррекции систематических ошибок прогноза без утечки данных в реальном времени. Экспериментальные результаты показывают, что стратегии консенсуса превосходят как отдельные модели, так и традиционное ансамблевое усреднение, а также дополнительные адаптивные бэйслайны. После коррекции систематических ошибок подход Skill-Weighted Voting достиг наивысшей общей точности с Kling-Gupta Efficiency 0.965 и Nash-Sutcliffe Efficiency 0.933, тогда как стратегия Adaptive Leader Selection оказалась наиболее робастной, обеспечив минимальную абсолютную ошибку 90-го перцентиля, тем самым снижая масштаб

крупных ошибок прогноза. Эти результаты демонстрируют, что сочетание консенсуса, вдохновленного блокчейном, с коррекцией ошибок в реальном времени является практичным и эффективным путем к созданию более устойчивых моделей прогнозирования.

Ключевые слова: прогнозирование стока, блокчейн, механизмы консенсуса, LSTM, коррекция смещения, прогнозирование наводнений, робастность моделей

Для цитирования: А.М. Альжанов, К.К. Рахымбек, А.Б. Нугуманова. Повышение устойчивости прогнозирования наводнений на основе ии с использованием консенсусных моделей, вдохновлённых блокчейном// Международный журнал информационных и коммуникационных технологий. 2025. Т. 6. No. 23. Стр. 24–44. (На анг.). https://doi.org/10.54309/IJICT.2025.23.3.002.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction

Accurate and reliable flood forecasting is a cornerstone of modern water resource management and disaster mitigation (Solanki et. al., 2025; Wee et. al., 2023). While artificial intelligence (AI) models have shown considerable promise in hydrology (Kratzert et. al., 2018. 6005–6022), their operational reliability can be compromised by noisy or anomalous input data, to which single models are particularly sensitive (Fang et. al., 2022). Ensemble forecasting offers a robust alternative, but conventional aggregation methods such as static averaging often fail to adapt to the dynamically changing performance of individual models, potentially dampening the signal from the most accurate model at any given time.

This study explores a more dynamic approach to ensemble forecasting by adapting principles from decentralized consensus mechanisms, conceptually like those used in blockchain technology (Umar et. al., 2025), focusing on the logic of dynamic, multi-agent agreement rather than a literal implementation of a distributed ledger or its cryptographic infrastructure. In this framework, individual models act as independent “nodes”, each proposing a forecast. Rather than being combined through a static aggregator, the forecasts are integrated using blockchain-inspired dynamic consensus strategies such as Quorum-based Agreement, Skill-Weighted Voting, and Adaptive Leader Selection.

A key aspect of this work is not only the application of these methods but also a detailed evaluation of their behavior and an investigation into their enhancement via post-processing. This evaluation includes an analysis of the internal behaviors of the consensus strategies, examining how the ensemble reaches agreement and adapts to the varying performance of individual models. Furthermore, we explore a potential enhancement for operational settings by augmenting the consensus output with an online, no-leakage debiasing technique to correct for systematic errors, with the goal of creating a more robust forecast.

Accordingly, this study is guided by the following research questions:

- RQ-1. Can blockchain-inspired consensus mechanisms improve the overall accuracy and robustness of streamflow forecasts compared to traditional ensemble averaging, single-model baselines and established adaptive ensemble methods?
- RQ-2. How do the different dynamic consensus strategies such as Quorum-based Agreement, Skill-Weighted Voting, and Adaptive Leader Selection compare in their performance and reliability?
- RQ-3. To what extent does applying an online, no-leakage debiasing technique to the consensus output further enhance forecast accuracy and reduce systematic errors?

This study puts forward a model-agnostic framework enhanced with real-time error correction, exploring its potential as a practical approach for operational streamflow forecasting systems. The findings are intended to demonstrate the benefits of adaptive aggregation in building more resilient hydrological prediction models.

Literature Review

Research at the intersection of blockchain and hydrology is still young but steadily growing. Two comprehensive reviews of distributed ledger applications in the water sector conclude that most implementations remain conceptual or lab-scale, with few operational deployments. Their findings emphasize that current efforts primarily target data integrity, auditability, and multi-party coordination rather than measurable improvements in forecast accuracy, while also highlighting unresolved issues such as interoperability and the need for reliable oracles to feed sensor data into on-chain logic (Satilmisoglu et. al., 2024; Asgari et. al., 2022).

Applications of blockchain in water systems generally fall into three categories. First, securing end-to-end data pipelines for flood detection and early warning has been explored through integration with UAV imagery, federated learning, and homomorphic encryption (Alsumayt et. al., 2023). Low latency alerting systems have also been tested, where smart contracts trigger warnings in lab settings (Wu et. al., 2024: 22–25), building on earlier conceptual visions such as the “Smart Dam” (Yasuno et. al., 2020: 139–158). Second, blockchain has been applied to ensure trustworthy ingestion and provenance of hydrometric data, for example through wireless sensor networks for pollution tracking (Lin et. al., 2020) or permissioned-chain frameworks for water quality monitoring (Le Thuy et. al., 2025; Vangipuram et. al., 2022). Third, automation of downstream actions such as insurance payouts and operational alerts has been studied in feasibility reports and prototype systems (Commonwealth Secretariat, 2022; Xia et. al., 2022). Collectively, these efforts emphasize blockchain as a back-end infrastructure for data security, integrity, and coordination rather than direct improvements to forecast accuracy.

In contrast, ensemble forecasting research in hydrology has primarily focused on improving predictive skill and robustness. Traditional approaches, such as regression-based model averaging (Williams et. al., 2016) and Bayesian Model Averaging (Torres et. al., 2024), adapt weights dynamically to changing model performance,

while newer paradigms such as consensus learning (Magureanu et. al., 2024) extend these ideas by allowing models to exchange predictions and reach agreement in the presence of faulty members. In meteorology, feature-oriented ensemble means have been developed for tropical cyclone forecasting, showing that adaptive aggregation can both improve track accuracy and better capture storm structure (Zhang et. al., 2021: 1945–1959). In hydrometeorological settings, deep learning is used not only to build base models but also as a meta-learner to combine model outputs or to post-process them, and in both cases, it often outperforms raw ensemble members and standard bias-correction baselines (Scher et. al., 2021; Dong et. al., 2025: 2023–2042).

Within hydrology, adaptive aggregation produces robust improvements across basins and lead times. A large sample Super Learner outperformed equal weighting and single models for daily streamflow (Tyralis et. al., 2021: 3053–3068). Bayesian Model Averaging has also been demonstrated in operational multi-reservoir inflow settings, highlighting its practical feasibility (Torres et. al., 2024). Dynamic, data-driven weighting, including time-series feature-based dynamic weights implemented within a BMA framework, improves merged forecasts across multiple lead times and flow regimes (Sheikh et. al., 2025: 1201–1217), which is consistent with a recent review advocating adaptive forecast-merging in practice (Sheikh et. al., 2024). Comparative assessments of bias correction and data assimilation further show that targeted post-processing can yield broad, reliable gains under operational constraints (Tanguy et. al., 2024: 1–41).

Current literature reveals a clear separation between applying blockchain for data integrity in hydrology and using adaptive ensembles to improve forecast accuracy. The critical gap lies at their intersection: the core consensus mechanisms that ensure blockchain's reliability have not been systematically adapted to enhance the skill of multi-model hydrological predictions.

Our study addresses this gap by applying blockchain-inspired consensus strategies in the Uba River Basin, augmenting them with an online no-leakage debiasing technique for operational robustness, and conducting an in-depth analysis of how these strategies function internally. This approach contributes both as a practical demonstration of consensus-inspired forecasting in a case study of Uba River basin and methodological insights into how such strategies can be enhanced for operational robustness.

Materials and Methods

Study Area and Data Description.

The Uba River, a 278 km-long right-bank tributary of the Irtysh River, originates from the confluence of the White and Black Uba rivers and flows through the East Kazakhstan. This study focuses on the 8,490 km² upstream catchment draining to the Shemonaikha gauging station (within the total 9,850 km² basin). The climate is sharply continental, with extreme temperatures ranging from −50 °C to +40 °C, and precipitation dominated by snowfall (>60 %), sustaining a persistent winter snowpack. The study area is illustrated in Figure 1.

In this study, we leverage two complementary sources of hydrometeorological data. First, we employ the Caravan dataset (Kratzert et. al., 2023), which provides daily streamflow observations from hundreds of catchments worldwide alongside basin-mean meteorological drivers derived from ERA5-Land (Muñoz-Sabater et. al., 2021: 4349–4383).

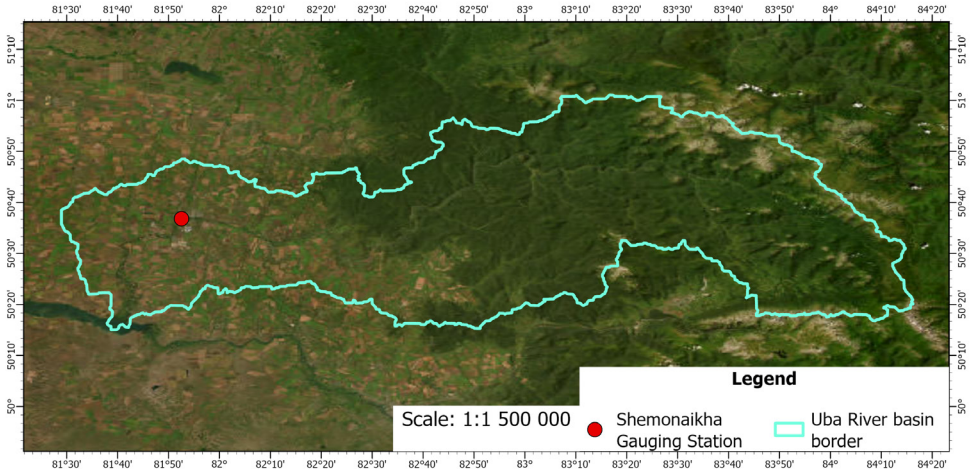


Fig. 1. Study area of Uba River basin with borders and the gauging station

These drivers include precipitation, temperature, evaporation, wind, soil moisture, and related variables aggregated across each basin. Second, we use ERA5-Land reanalysis data directly for the Uba River Basin, extracting predictors at multiple spatial scales: (i) mean values across the basin, (ii) values from the exact grid cell containing the Shemonaikha gauge, and (iii) the full grid of values covering the catchment. The target variable is daily streamflow observed at the Shemonaikha gauging station (50.61°N, 81.87°E).

Baseline Models

This study evaluates four LSTM-based models for daily streamflow forecasting in the Uba River basin, each differing in spatial input structure and training scope. The models are:

- LSTM-L: A lumped, single-basin model trained only on the Uba basin.
- LSTM-Caravan: A lumped, multi-basin model trained on 150 globally distributed basins, including Uba.
- LSTM-Grid: A gridded model trained on inputs from all grid cells across the Uba basin.
- LSTM-1-Cell: A single-point model trained on input from only the stream gauge's grid cell.

All models were trained on data from 1995–2009 and validated on 2010–2011. To ensure a fair comparison, the testing period of 2012–2020 used consistent hydrological-year splits across all experiments. Table 1 summarizes the key characteristics of the four LSTM models used in this study.

These four models serve as the foundational components from which the final

ensemble predictions are derived. Their varied spatial structures and training scopes provide a robust and diverse set of inputs for the blockchain-inspired consensus strategies used in this study.

Table 1. Summary of LSTM-Based Baseline Models for Streamflow Forecasting in the Uba River Basin

Model	Spatial setup	Training scope	Dynamic inputs	Static inputs	Seasonal scope	Architecture and training configurations
LSTM-L	Basin-averaged (lumped)	Uba basin only	39 daily ERA5-Land variables aggregated to basin mean	210 basin attributes	Full year	LSTM(256); Dropout: 0.4; Optimizer: Adam; LR: 5×10^{-4}; Epochs: 50
LSTM-Caravan		150 Caravan basins + Uba basin				
LSTM-Grid	Gridded	Uba basin (full spatial extent)	Daily mean/max air temp, precipitation, SWE, soil temp (0-7 cm), soil moisture (0-7 cm), per grid cell	Latitude/ Longitude per cell	Nov-May	2-layer LSTM(65); Dropout: 0.2; Optimizer: AdamW; LR: 5×10^{-4}; Batch: 64; Epochs: 50 (best validation checkpoint)
LSTM-1-Cell	Single grid cell	Uba basin (gauging cell only)	Same as LSTM-Grid, but only for the gauging cell			

Ensemble Baselines

In addition to the four individual LSTM models, three ensemble-based methods were evaluated to provide a broader comparative framework. These include one traditional static approach and two adaptive techniques that operate exclusively on the forecast outputs of the LSTM models. All ensemble methods were applied during the testing period (2012–2020) and do not involve access to raw meteorological inputs or modification of the underlying model parameters.

The Ensemble-Avg (a simple arithmetic mean of the four models), represents the traditional ensemble approach. This comprehensive suite of baselines allows for a direct comparison against both high-performing individual models and the standard approach for ensemble forecasting.

The Online Super Learner (OSL) is an adaptive aggregation method that estimates a weighted combination of model forecasts using ridge-regularized linear regression over a rolling window of past observations. The regression is fit without an intercept and is re-fit on each rolling window. At each time step, weights are updated by minimizing the regularized prediction error on a recent history of forecast-observation pairs. An exponential forgetting factor is applied as sample weights on the

windowed observations to emphasize recent data, and the regression coefficients are subject to non-negativity and unit-sum constraints, enforced through projection onto the probability simplex.

Dynamic Model Averaging (DMA) extends the Bayesian model averaging by allowing weights to adapt over time. Model weights are recursively updated at each time step based on the predictive likelihood of recent residuals. A forgetting factor (applied via elementwise exponentiation) controls memory decay, and model likelihoods (raised to a temperature power) determine the sensitivity of the subsequent weight update.

Blockchain Consensus-Inspired Ensemble Strategies

A series of consensus-inspired ensemble strategies was designed to combine predictions from the baseline models, thereby enhancing forecasting accuracy and reliability. This study examines three alternative consensus strategies:

- Quorum-based Median: A median forecast derived from a quorum of predictions that fall within a defined tolerance envelope.
- Skill-Weighted Voting: A weighted average where each model's prediction is weighted based on its historical performance.
- Adaptive Leader Selection: An adaptive approach where a single leader is selected based on a rolling evaluation of performance. A robust fallback mechanism is included to ensure stability.

Quorum-based Median. The Quorum-based Median agreement strategy, inspired by Byzantine fault tolerance (BFT) principles, produces a consensus forecast only if enough models agree within the defined tolerance envelope. The initial step involves determining the median () at time from all baseline model outputs (), which serves as the central reference point for evaluating agreement. To assess the level of agreement among predictions, a tolerance envelope is constructed using both absolute and relative criteria. The absolute deviation is quantified using the Median Absolute Deviation (MAD), denoted as , which is the median of the absolute differences across all models. This provides a robust measure of dispersion that is less influenced by outliers than the standard deviation.

The tolerance threshold, denoted τ_t, is then defined by combining the MAD and the magnitude of the ensemble median, as shown below:

$$\tau_t = \max\left(\alpha \cdot MAD_t, \beta \cdot |m_t|\right), \quad (1)$$

where α and β are predefined hyperparameters controlling the sensitivity to deviations in absolute and relative terms, respectively. A model prediction $x_{t,i}$ agrees if it falls within this dynamic envelope around the median. The final consensus forecast ($\widehat{y_t}$), is determined by a quorum condition. If the number of agreeing models is at least $\lceil q \cdot M \rceil$, where q is the quorum fraction and M is the number of models, the consensus is defined as the median of agreeing predictions:

$$\hat{y}_t = \text{median}\{x_{t,i} : |x_{t,i} - m_t| \leq \tau_t\} \quad (2)$$

If the quorum is not met, the fallback is the overall ensemble median (m_t), which ensures a stable output even during periods of low agreement.

Skill-Weighted Voting. The Skill-Weighted Voting method, analogous to Proof-of-Stake (PoS) protocols, constructs the consensus forecast as a weighted average of individual model predictions. Each model i is assigned a weight based on its historical skill score (S_i), evaluated with a performance metric. To avoid zero or negative weights, the scores are floored by a small constant ε. The normalized weights are then defined as:

$$w_i = \frac{(\max(S_i,0)+\varepsilon)^p}{\sum_{j=1}^{M}(\max(S_j,0)+\varepsilon)^p}, \quad (3)$$

where p is the skill power hyperparameter that controls how strongly high-skill models are emphasized. The consensus forecast at time t is then computed as the convex combination of the model forecasts:

$$\hat{y}_t = \sum_{i=1}^{M} w_i \, x_{i,t} \quad (4)$$

Adaptive Leader Selection. The Adaptive Leader Selection is a strategy, like Delegated Proof-of-Stake (DPoS), that dynamically delegates forecasting authority to the model with the highest rolling skill. This approach leverages the strengths of the most reliable model while reducing the risk of persistent failures through a robust fallback mechanism. The leader (L_t), is the model that has demonstrated the highest skill over a defined rolling window ($W_l(t)$). The skill is determined by minimizing a given error metric. The leader is selected according to the following equation:

$$L_t = arg\min_i S_{t,i}(w), \quad (5)$$

where $S_{t,i}(w)$ is the skill score of the model i at time t. The consensus forecast is then simply the prediction of the leader:

$$\hat{y}_t = x_{t,L_t} \quad (6)$$

To avoid unstable switching and ensure temporal smoothness, a veto check is applied. If the leader's forecast deviates from the ensemble median m_t beyond the tolerance envelope τ_t, it is vetoed. The condition for this check is as followed:

$$\hat{y}_t |x_{t,L_t} - m_t| > \tau_t, \quad (7)$$

Here, τ_t is the tolerance defined in the Quorum-based Median strategy (Equation 1). If the condition in Equation (7) is met, the method falls back to a more stable consensus, such as the overall ensemble median or the median of the agreeing set.

This robust safety check ensures that the system does not rely on a single, potentially incorrect, prediction, thereby providing a resilient forecasting strategy.

Experimental Design

The experimental design was structured to systematically address the study's research questions by evaluating the performance of the consensus-inspired strategies against the pre-defined baselines, both with and without a post-processing debiasing module. Our approach is conceptually inspired by the architecture of a blockchain, as illustrated in Figure 2. In this framework, each of the four LSTM models is treated as an independent "node" that provides a prediction for a specific time step. These predictions are aggregated within a sequential "block" which contains the consensus mechanism and bias adjustment layers to produce a final forecast for this time step. These blocks are linked chronologically to form a chain of forecasts.

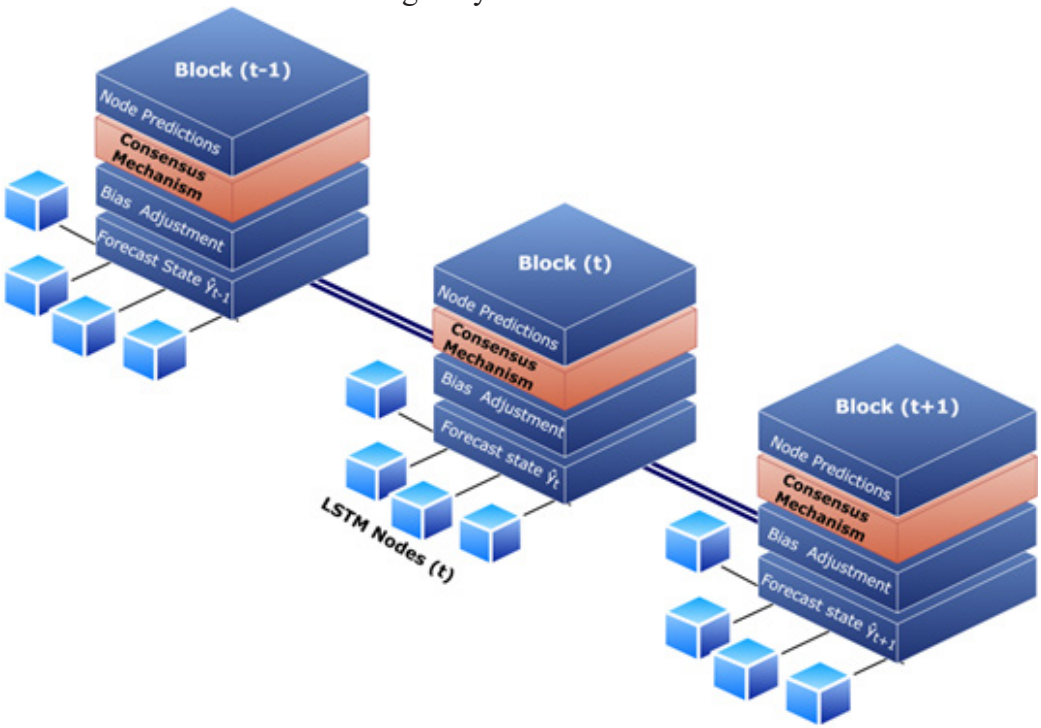


Fig. 2. Conceptual framework of the blockchain-inspired consensus forecasting model

While our approach is conceptually inspired by the architecture of a blockchain, it is important to clarify the limits of this analogy. We adopt the terminology of "nodes" (individual models) and sequential "blocks" (time-step forecasts) to describe the information flow, as illustrated in Fig. 2. However, our centralized framework does not involve cryptographic hashing, a distributed peer-to-peer network, or immutable ledgers. The analogy serves to highlight the process of reaching a reliable outcome from multiple, independent, and potentially conflicting sources, which is the core principle we borrow from blockchain's consensus mechanisms. To test this framework, the experiment was conducted in two sequential stages, and the outputs of each stage were evaluated.

In the first stage, the daily streamflow predictions from the four baseline LSTM models and the Ensemble-Avg were used as inputs. These forecasts were processed by the three consensus-inspired strategies described in this study. Each strategy was applied in a post-training fashion, meaning the internal structure and weights of the baseline models remained unchanged. The consensus mechanisms relied solely on the ensemble's outputs and, for skill-aware strategies, on historical performance statistics. This stage directly addresses RQ-1 and RQ-2 by allowing for a direct comparison of the consensus models' performance against both the individual models and ensemble average baseline.

In the second stage, a post-processing debiasing module was applied to correct systematic errors in the forecasts produced by all individual models, ensemble average baseline and consensus strategies. This was done to assess the extent to which an online, no-leakage error correction could enhance operational reliability, thereby addressing RQ-3. The bias was computed as the rolling mean of recent prediction errors (observed minus predicted streamflow). The corrected forecast was then obtained through an additive adjustment, as shown:

$$\hat{y}_t^{corr} = y_t + mean(o_{t-W_b+1:t} - y_{t-W_b+1:t}) \quad (8)$$

where $\hat{y}_t^{corr}$ is the bias corrected forecast, y_t is the original model prediction, o_t is the observed streamflow, and W_b is the length of the debiasing window used to calculate the rolling mean error.

Each experiment was executed using the same set of input predictions from the baseline models, allowing for a direct comparison under identical conditions.

Sensitivity analysis

To evaluate the influence of key hyperparameters and support the selection of final configurations, a sensitivity analysis was conducted to evaluate the effect of key hyperparameters across ensemble methods, using KGE as the performance metric, as shown in Figure 3.

The sensitivity analysis revealed three key patterns:

- Temporal adaptivity. Parameters controlling memory and responsiveness had the strongest impact. Shorter window lengths improved responsiveness across methods. Optimal values were 10 days for debiasing/adaptive leader selection, 30 days for OSL, and 90 days for DMA. Forgetting factors were equally important: OSL achieved its best performance at 0.94, and DMA at 0.992, indicating a preference for recent observations in dynamic weighting.
- Weighting schemes. Strategies relying on selective model weighting showed clear trends. In Quorum-based Median, performance improved with increasing agreement threshold, reaching its maximum at $q = 1.0$. In Skill-Weighted Voting, accuracy increased with stronger emphasis on top-performing models, peaking at a skill power of $p = 20$.
- Stability parameters. Parameters intended for regularization or smoothing had

limited influence. OSL's ridge penalty (λ_2) had negligible effect across tested values, DMA's likelihood temperature provided only marginal gains, with best performance at 1.0.

Evaluation

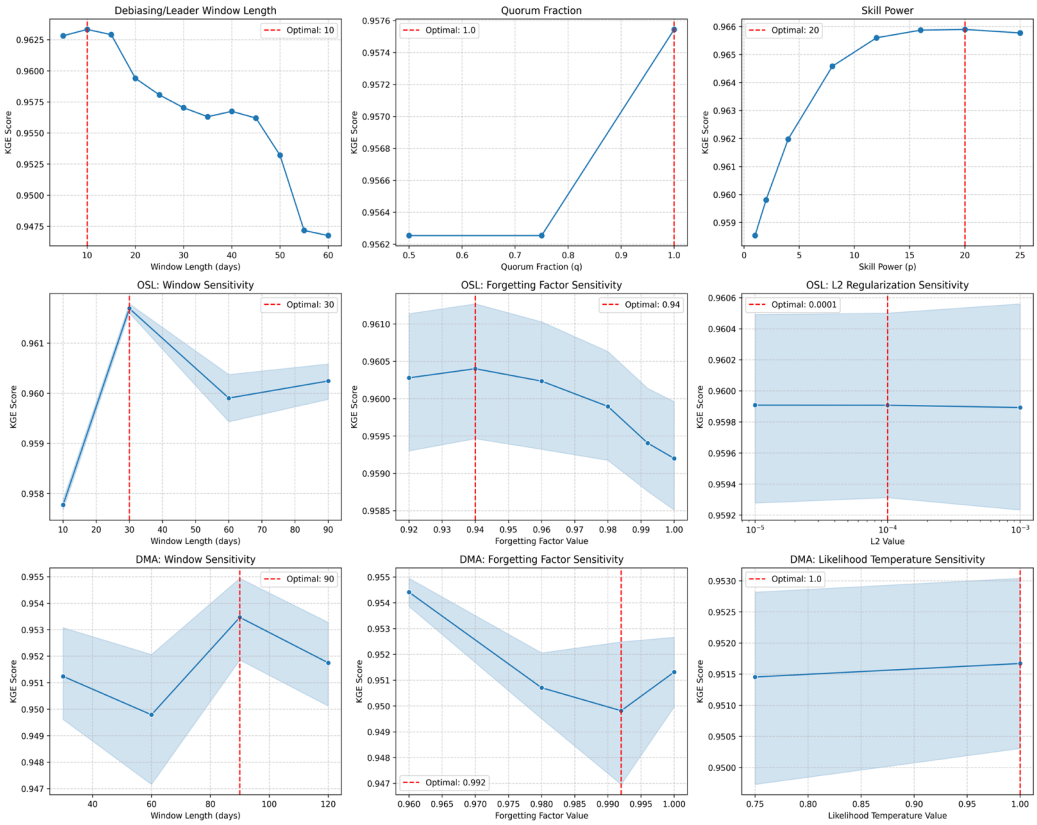


Fig. 3. KGE-based sensitivity analysis of key hyperparameters across all evaluated methods, red dashed lines indicate optimal values selected for final experiments

Predictive performance was evaluated using a set of standard hydrological metrics, each capturing different aspects of forecasting capability. The Nash-Sutcliffe Efficiency (NSE) was used to assess overall accuracy, where a value of 1 indicates perfect agreement and 0 reflects performance equivalent to the observed mean. It is defined as:

$$NSE = 1 - \frac{\sum_{t=1}^{n}(Q_{sim,t} - Q_{obs,t})^2}{\sum_{t=1}^{n}(Q_{obs,t} - \bar{Q}_{obs})^2} \quad (9)$$

where $Q_{sim,t}$ is the simulated streamflow at time t, $Q_{obs,t}$ is the observed streamflow, and $\bar{Q}_{obs}$ is the mean observed streamflow over the evaluation period.

The Kling–Gupta Efficiency (KGE) was applied to provide a diagnostic view of performance by combining correlation (r), bias ratio (β), and variability ratio (α):

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\alpha-1)^2} \quad (10)$$

Systematic errors across the hydrograph were further assessed using Flow Duration Curve (FDC) biases, focusing on two segments: High-Flow Bias (FHV), which captures errors in the top 2 % of flows (peak events), and Low-Flow Bias (FLV), which reflects performance in the lowest 30 % of flows (baseflow conditions).

Results

This section presents a detailed evaluation of the forecasting models, structured to directly address the study's research questions. To answer RQ-1 and RQ-2, we first compare the overall predictive accuracy of the blockchain-inspired consensus mechanisms against individual models and ensemble baselines. Subsequently, to address RQ-3, we assess the enhancement provided by the online debiasing technique, focusing on improvements in overall accuracy and model reliability by analyzing the 90th percentile of their prediction errors. A comprehensive summary of all performance metrics is provided in Table 2 for reference.

Table 2. Summary of Key Performance Metrics for All Configurations

Model Configuration	KGE	NSE	FHV (%)	FLV (%)
Configurations with debias				
Skill-Weighted Voting + Debias	0.9658	0.9337	-9.23	-104.87
Adaptive Leader Selection + Debias	0.9633	0.9267	-11.42	516.34
LSTM-Grid + Debias	0.9625	0.9287	-10.52	-65.18
Online Super Learner + Debias	0.9619	0.9279	-11.98	-35.76
Quorum-based Median Agreement + Debias	0.9575	0.9306	-14.28	-134.57
Ensemble-Avg + Debias	0.9571	0.9326	-13.27	-92.03
Dynamic Model Averaging + Debias	0.9564	0.9129	-7.3	-134.56
LSTM-Caravan + Debias	0.954	0.9101	-7.58	-134.56
LSTM-1-Cell + Debias	0.9368	0.8783	-16.35	84.37
LSTM-L + Debias	0.9361	0.8829	-18.6	-138.46
Configurations without debias				
Dynamic Model Averaging	0.9428	0.8866	-13.64	-879.22
Skill-Weighted Voting	0.9393	0.9081	-17.41	-997.3
LSTM-Caravan	0.9304	0.8759	-13.64	-879.22
Adaptive Leader Selection	0.9299	0.9093	19.5	-321.48
LSTM-Grid	0.9198	0.8997	-20.86	-1214.52
Online Super Learner	0.8623	0.8954	-25.89	-560.38
Quorum-based Median Agreement	0.8605	0.8962	-27.5	-525.69
Ensemble-Avg	0.857	0.8985	-26.37	-541.48
LSTM-L	0.8103	0.8106	-32.7	-172.16
LSTM-1-Cell	0.6884	0.7941	-38.28	100

Overall Performance Analysis

The initial performance of the non-debiased models revealed significant variability and systemic biases. Among them, Dynamic Model Averaging achieved the

highest non-debiased KGE (0.943), while Adaptive Leader Selection produced the best NSE (0.909). However, all original models struggled profoundly with low flows, exhibiting severe underestimation biases (FLV). This highlighted a fundamental weakness in the models' ability to capture the full range of hydrological conditions.

The application of an online, no-leakage debiasing technique leads to a consistent improvement in overall accuracy across all models, as shown in Figures 4 and 5. With debiasing applied, the Skill-Weighted Voting approach achieved the highest NSE (0.933) and KGE (0.965), while other adaptive ensemble approaches and individual models also benefited. These results demonstrate that while blockchain-inspired consensus methods offer an advantage, their combination with an online error correction mechanism is essential for achieving robust and superior forecast accuracy.

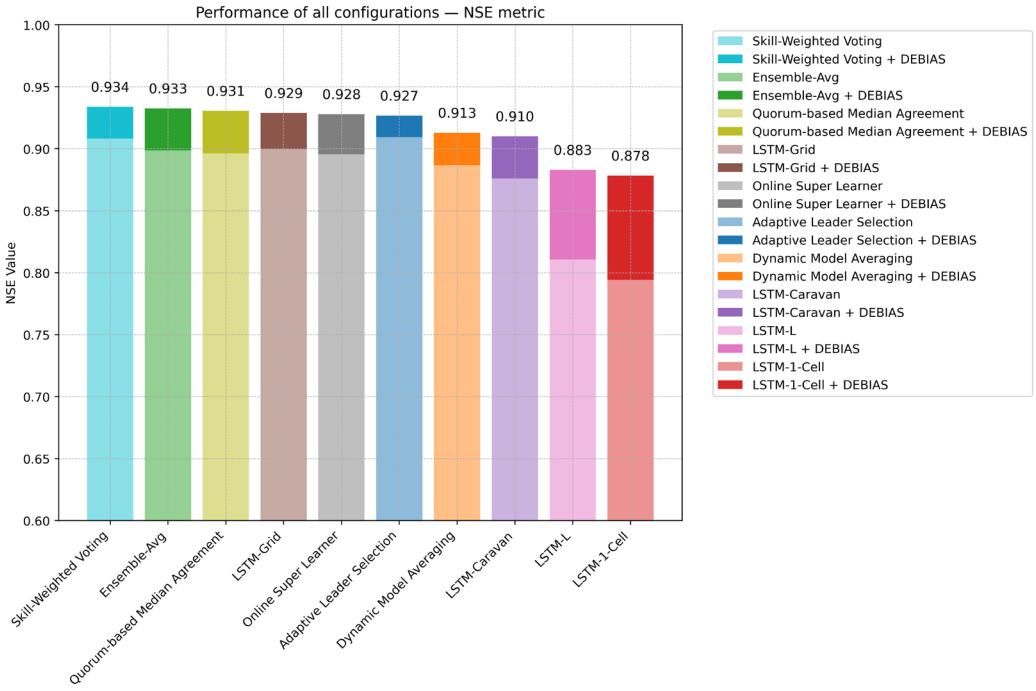


Fig. 4. NSE Performance Across All Configurations

Forecast Reliability and Robustness

To assess model reliability and robustness, we analyzed the 90th percentile of absolute errors (Q90 Absolute Error), which represents the average error of the worst 10 % of predictions. This metric is important for operational forecasting, as it quantifies a model's tendency for making large, potentially misleading errors.

As shown in Figure 6, the Adaptive Leader Selection + Debias strategy produced the lowest error in these worst-case scenarios, indicating it is the least likely to produce extreme forecast errors. In contrast, the individual LSTM models exhibited the largest errors, highlighting their unreliability in comparison to adaptive ensemble approaches which are more dependable by minimizing the magnitude of their largest errors.

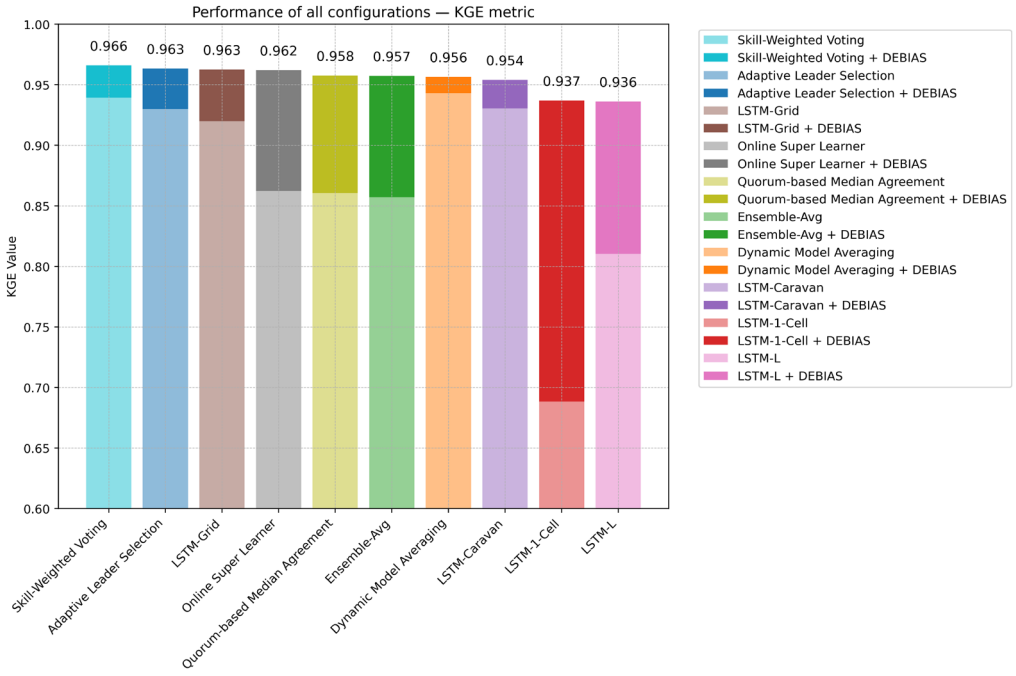


Fig. 5. KGE Performance Across All Configurations

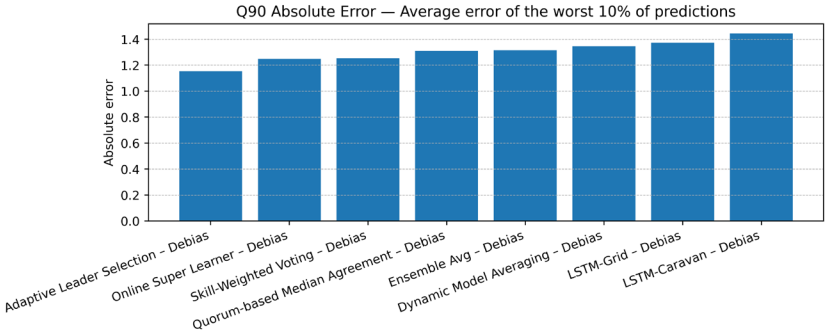


Fig. 6. Average Error of the Worst 10% of Predictions

Internal Dynamics of Adaptive Leader Selection

Although Skill-Weighted Voting after debiasing delivered the best overall accuracy in terms of KGE and NSE, Adaptive Leader Selection stood out for its ability to control extreme errors, achieving the lowest Q90 absolute error (Figure 6). To better understand the mechanism behind it, we examined its internal behavior. Figure 7 illustrates which of the four baseline models was selected as the “leader” at each time step throughout the test period.

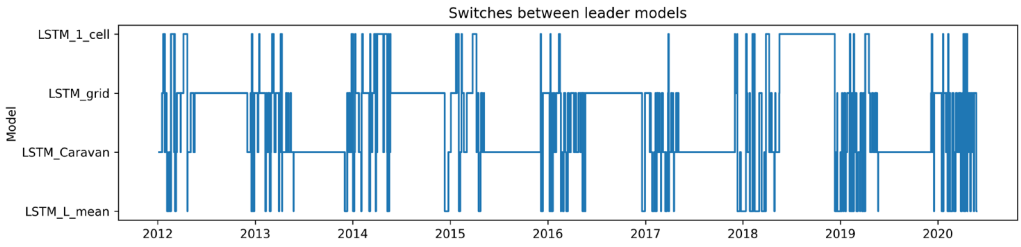
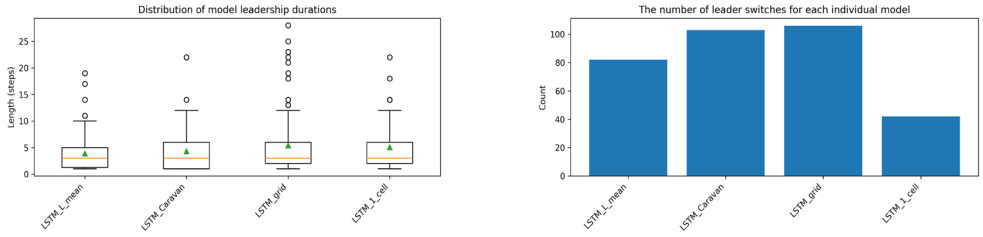


Fig. 7. Leader Model Selection Over Time

While Figure 7 visually shows that the strategy frequently switches its reliance between the different LSTM configurations, Figure 8 quantifies these dynamics by showing the distribution of leadership durations lengths and number of leader switches for each model. It shows that LSTM-Caravan and LSTM-Grid were the most frequently chosen leaders, with other models selected less often, reflecting their more limited role in the ensemble. Further analysis of leadership duration distributions highlights that LSTM-Grid not only switched frequently but also sustained leadership for longer periods, as indicated by its wider interquartile range and multiple long-duration outliers.

This quantitative evidence complements the visual analysis from Figure 7, suggesting the strategy's success may be attributed to its ability to rely on generally robust models as a default, while maintaining the agility to switch to more specialized models for short bursts when conditions demand it.



(a) *(b)*

Fig.8. Quantifying the Internal Dynamics: (A) Distribution of Model Leadership Durations Lengths (B) the Number of Leader Switches for Each Model

In summary, the results demonstrate that blockchain-inspired consensus models, particularly when enhanced with a debiasing technique, provide an improvement over both individual models and traditional ensemble averaging. Skill-Weighted Voting achieved the highest overall accuracy, while Adaptive Leader Selection provided the best control of extreme errors, highlighting their complementary strengths.

Discussion

This section delves into the interpretation of the results, directly addressing the research questions set out in the introduction and exploring the practical implications of the findings.

In response to our first research question (RQ-1), the results confirm that

blockchain-inspired consensus models improve both the accuracy and robustness of streamflow forecasts. As a top-performing approach, Skill-Weighted Voting demonstrated superior performance in both KGE and NSE against the standard Ensemble-Avg, the individual LSTM models and alternative adaptive ensemble methods. These findings validate the application of blockchain-inspired consensus principles as an effective approach for dynamic model aggregation in hydrology.

Regarding the comparison of different consensus strategies (RQ-2), Skill-Weighted Voting provided the most well-rounded performance, attaining the highest overall NSE and KGE. Adaptive Leader Selection proved to be the most robust, achieving the lowest 90th percentile error, making it particularly suited for operational scenarios where minimizing large errors is critical. An internal analysis of the Adaptive Leader Selection strategy revealed that its robustness to large errors stems from its ability to identify consistently strong models as primary leaders, while still leveraging the strengths of other ensemble members by switching to them as hydrological conditions change.

Finally, addressing our third research question (RQ-3), the study found that applying an online, no-leakage debiasing technique substantially improved the operational reliability of the forecasts. This post-processing step produced significant performance gains across all models and consensus strategies. Most importantly, it corrected the severe systemic biases, particularly the underestimation of low flows, that were prevalent in the original models, thereby enhancing their robustness for operational use.

It is important to note that the proposed framework operates entirely as a post-processing layer on top of daily forecasts already produced by the baseline LSTM models. As such, the consensus strategies only require combining a handful of scalar predictions at each time step. This renders the computational overhead negligible relative to the cost of model inference and data preparation. Since the framework is model-agnostic and designed to work with any set of base forecasts, its integration into an operational setting does not introduce latency or scalability concerns. For this reason, we did not provide a detailed runtime analysis, as computational feasibility is not a limiting factor for the proposed approaches.

Conclusion

This study investigated the potential of blockchain-inspired consensus models, augmented with an online debiasing technique, to improve the accuracy and reliability of streamflow forecasting in the Uba River Basin. Through this evaluation, we found that the blockchain-inspired dynamic aggregation strategies employed in this study outperform the baseline approaches of traditional static averaging, individual models, and other established ensemble strategies such as Dynamic Model Averaging and Online Super Learner.

The primary implication of this work is that principles inspired by decentralized consensus in blockchain technology offer a valuable framework for improving hydrological forecasting. A key strength of this model-agnostic framework is its flex-

ibility, it can be applied to any collection of forecasting models, regardless of their internal architecture. Combining these consensus-driven aggregation methods with real-time error correction can contribute to creating forecasting systems that are more accurate, robust, and resilient to the inherent uncertainties of hydrological processes.

Future work should focus on validating the generalizability of this blockchain-inspired consensus framework across a broader range of catchments with diverse hydrological and climatic regimes. Extending the approach to ensembles that include not only data-driven models but also process-based hydrological models could provide further enhance robustness under varying conditions. In addition, exploring a wider range of consensus mechanisms such as hierarchical, trust-based, or hybrid strategies has the potential to clarify which approaches are most effective for hydrological forecasting. Extending the framework to multi-basin applications could involve developing spatially aware skill metrics, where a model's credibility in one catchment is informed by its performance in hydrologically similar or adjacent regions. Aside from that, the regional models trained on multiple basins using dynamic and static features could be utilized to form ensemble for this framework. For deployment in real-time operational systems, a crucial next step is to integrate the consensus module with live data assimilation workflows, enabling skill weights to update immediately as new observations become available.

Funding. This work has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant Number BR24992899).

REFERENCES

Asgari M., Nemati M. (2022). Application of distributed ledger platforms in smart water systems — A literature review // Frontiers in Water. — 2022. — Vol. 4. — Article 848686. — DOI: 10.3389/frwa.2022.848686.

Alsumayt A., El-Haggar N., Amouri L., Alfawaer Z.M., Aljameel S.S. (2023). Smart flood detection with AI and blockchain integration in Saudi Arabia using drones // Sensors. — 2023. — Vol. 23, No. 11. — Article 5148. — DOI: 10.3390/s23115148.

Commonwealth Secretariat. Parametric climate insurance using blockchain technology: A feasibility study on technical and economic viability in Fiji. — London: Commonwealth Secretariat, 2022. — (Report). — DOI: 10.14217/comsec.1087.

Dong N., Hao H., Yang M., Wei J., Xu S., Kunstmann H. (2025). Deep-learning-based sub-seasonal precipitation and streamflow ensemble forecasting over the source region of the Yangtze River // Hydrology and Earth System Sciences. — 2025. — Vol. 29, No. 8. — pp. 2023–2042. — DOI: 10.5194/hess-29-2023-2025.

Fang K., Kifer D., Lawson K., Feng D., Shen C. (2022). The data synergy effects of time-series deep learning models in hydrology // Water Resources Research. — 2022. — Vol. 58, No. 4. — Article e2021WR029583. — DOI: 10.1029/2021WR029583.

Kratzert F., Nearing G., Addor N., Erickson T., Gauch M., Gilon O., et al. (2023). Caravan — A global community dataset for large-sample hydrology // Scientific Data. — 2023. — Vol. 10. — No. 1. — Article 61. — DOI: 10.1038/s41597-023-01975-w.

Muñoz-Sabater J., Dutra E., Agustí-Panareda A., Albergel C., Arduini G., Balsamo G., et al. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications // Earth System Science Data. — 2021. — Vol. 13. — No. 9. — Pp. 4349–4383. — DOI: 10.5194/essd-13-4349-2021.

Kratzert F., Klotz D., Brenner C., Schulz K., Herrnegger M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks // Hydrology and Earth System Sciences. — 2018. — Vol. 22. — No. 11. — Pp. 6005–6022. — DOI: 10.5194/hess-22-6005-2018.

Lin Y.P., Mukhtar H., Huang K.T., Petway J.R., Lin C.M., Chou C.F., Liao S.W. (2020). Real-time identifica-

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Правила оформления статьи для публикации в журнале на сайте:
https://journal.iitu.edu.kz

ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Собственник: АО «Международный университет информационных технологий» (Казахстан, Алматы)

Подписано в печать 15.09.2025.
Формат 60x881/8. Бумага офсетная. Печать - ризограф. 9,0 п.л. Тираж 100
050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09).

Издание Международного университета информационных технологий
Издательский центр КБТУ, Алматы, ул. Толе би, 59