Ko‘p maqsadli gidroenergetika majmualarida tog‘ daryolari gidroenergetik resurslarini uzoq muddatli prognozlashga mo‘ljallangan GIS asosidagi yirik ma’lumotlarni qayta ishlash tizimi

Nurbek N. Kurbonov [1,a)], Shohruh M. Atajiyev[1], Shohruh S. Samiyev[1,b)]

[1,a)] PhD, Toshkent davlat texnika universiteti, Tashkent; nurbek.kurbonov.96@gmail.com http://orcid.org/0009-0006-5400-3532

[1]Tayanch doktorant, Toshkent davlat texnika universiteti, Toshkent, 100000, O‘zbekiston; mr.shoh.mr@gmail.com https://orcid.org/0009-0007-1833-0461

[1,b)]Assistent, Toshkent davlat texnika universiteti, Toshkent, 100000, O‘zbekiston; mr.shoh.mr@gmail.com https://orcid.org/0009-0007-1833-0461

Dolzarblik: Gidroenergetika resurslarini uzoq muddatli prognozlash uchun katta hajmdagi gidrologik va geofazoviy ma’lumotlarni intellektual qayta ishlash zarur.

Maqsad: Ushbu tadqiqotda ko‘p maqsadli gidroenergetika majmualari uchun GIS asosidagi Big Data processing tizimi ishlab chiqilib, uzoq muddatli prognozlash aniqligini oshirish maqsadida taqsimlangan hisoblash, statistik qayta ishlash, IoT monitoring va geoinformatsion texnologiyalar integratsiyasi amalga oshirildi.

Usullar: PySpark asosidagi taqsimlangan hisoblash, Gauss filtrlash, log-normal transformatsiya, Puasson va binomial statistik qayta ishlash usullari qo‘llanildi.

Natijalar: Taklif etilgan tizim ma’lumotlar ishonchliligini 0,96–0,99 gacha oshirdi, anomal qiymatlarni kamaytirdi va uzoq muddatli prognozlash barqarorligini yaxshiladi.

Kalit so‘zlar: Big Data processing; GIS; gidroenergetika prognozi; PySpark; gidrologik modellashtirish; IoT monitoring; taqsimlangan hisoblash; tog‘ daryolari.

ГИС-ориентированная система обработки больших данных для долгосрочного прогнозирования гидроэнергетических ресурсов горных рек в многоцелевых гидроэнергетических комплексах

Нурбек Н. Курбонов[1,a)], Шохрух М. Атажиев[1], Шохрух С. Самиев[1,b)]

[1,a)] PhD, Ташкентский государственный технический университет; nurbek.kurbonov.96@gmail.com http://orcid.org/0009-0006-5400-3532

[1]Докторант, Ташкентский государственный технический университет, Ташкент, 100000, Узбекистан; mr.shoh.mr@gmail.com https://orcid.org/0009-0007-1833-0461

[1,b)]Ассистент, Ташкентский государственный технический университет, Ташкент, 100000, Узбекистан; mr.shoh.mr@gmail.com https://orcid.org/0009-0007-1833-0461

Актуальность: Точное долгосрочное прогнозирование гидроэнергетических ресурсов требует интеллектуальной обработки больших гидрологических и геопространственных данных.

Цель: В работе разработана GIS-ориентированная система обработки Big Data для долгосрочного прогнозирования гидроэнергетических ресурсов горных рек с использованием распределённых вычислений, статистической предобработки, IoT-мониторинга и геоинформационных технологий для повышения точности прогнозирования и надёжности гидроэнергетических комплексов.

Методы: Применялись распределённая обработка на основе PySpark, гауссовская фильтрация, лог-нормальное преобразование, распределения Пуассона и биномиальная статистическая обработка.

Результаты: Предложенная система повысила надёжность данных до 0,96–0,99, уменьшила аномалии и улучшила устойчивость долгосрочного прогнозирования гидроэнергетических ресурсов.

Ключевые слова: обработка Big Data; GIS; прогнозирование гидроэнергетики; PySpark; гидрологическое моделирование; IoT-мониторинг; распределённые вычисления; горные реки.

For citation: Kurbonov N.N., Shokhrukh Atajiev Sh.M., Samiev Sh.S. Gis-based big data processing framework forlong-term forecasting of mountain river hydropower resources in multi-purpose hydropower complexes. Scientific and technical journal of Problems of Energy and Sources Saving, 2026, no. 2, pp. 127-131.

https://doi.org/10.5281/zenodo.21205690

Received: 18.12.2025
Revised: 17.04.2026
Accepted: 10.06.2026
Published: 26.06.2026

Gis-based big data processing framework forlong-term forecasting of mountain river hydropower resources in multi-purpose hydropower complexes

Nurbek N. Kurbonov [1,a], Shokhrukh M. Atajiev[1], Shokhrukh S. Samiev[1,b]

[1,a] PhD, Tashkent State Technical University, Tashkent; nurbek.kurbonov.96@gmail.com http://orcid.org/0009-0006-5400-3532
[1]PhD student, Tashkent State Technical University, Tashkent, 100000, Uzbekistan; mr.shoh.mr@gmail.com https://orcid.org/0009-0007-1833-0461
[1,b]Assistant, Tashkent State Technical University, Tashkent, 100000, Uzbekistan; mr.shoh.mr@gmail.com https://orcid.org/0009-0007-1833-0461

Relevance: Accurate long-term hydropower forecasting requires intelligent Big Data processing of heterogeneous hydrological and geospatial datasets.
Objective: This study develops a GIS-based Big Data processing framework for long-term forecasting of mountain river hydropower resources using distributed computing, statistical preprocessing, IoT monitoring, and geospatial integration technologies to improve forecasting accuracy and operational reliability of multi-purpose hydropower complexes.
Methods: PySpark-based distributed processing, Gaussian filtering, log-normal transformation, Poisson and binomial statistical preprocessing methods were applied.
Results: The proposed framework improved data reliability to 0,96–0,99, reduced anomalies, and enhanced long-term hydropower forecasting stability and processing efficiency.
Keywords: Big Data processing; GIS; hydropower forecasting; PySpark; hydrological modeling; IoT monitoring; distributed computing; mountain rivers.

1. Introduction

The rapid growth of renewable energy systems and increasing electricity demand require the development of intelligent forecasting systems for sustainable hydropower resource management. Hydropower remains one of the most reliable renewable energy sources because of its operational flexibility, energy storage capability, and long-term stability. In mountainous regions, multi-purpose hydropower complexes simultaneously perform electricity generation, irrigation regulation, flood protection, and water resource balancing functions. Therefore, accurate long-term forecasting of mountain river hydropower potential becomes an important scientific and engineering problem. Hydrological systems of mountainous river basins are characterized by nonlinear and highly dynamic processes influenced by precipitation variability, glacier melting, snow cover changes, seasonal temperature fluctuations, and reservoir operational conditions. Traditional statistical approaches are insufficient for analyzing continuously growing hydrological and geospatial datasets generated by modern monitoring systems. Consequently, the integration of Big Data technologies, GIS platforms, SCADA systems, and IoT-based monitoring architectures becomes necessary. Modern distributed computing technologies such as Hadoop and Spark significantly improve the capability for processing large-scale hydrological information. At the same time, machine learning and hybrid hydrological forecasting models demonstrate high efficiency for river flow and reservoir operation prediction. These approaches create the basis for intelligent forecasting systems capable of supporting sustainable operation of hydropower complexes [1-3].

The proposed study focuses on the development of a GIS-based Big Data processing framework for long-term forecasting of mountain river hydropower resources in multi-purpose hydropower complexes. The proposed architecture integrates IoT monitoring technologies, SCADA systems, distributed storage environments, geospatial information systems, and statistical preprocessing approaches for improving forecasting reliability and operational efficiency (Fig. 1).

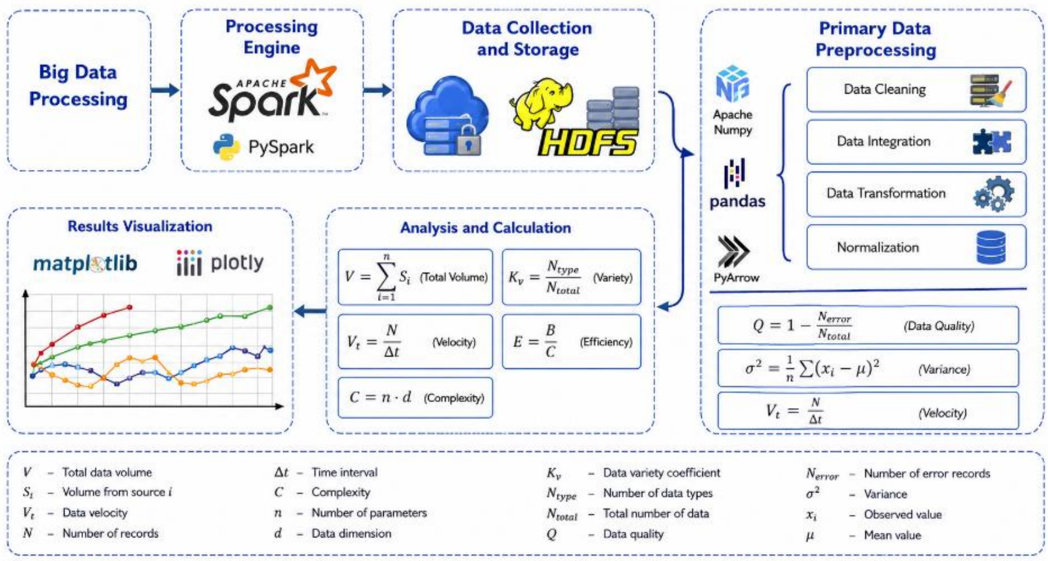

Fig. 1. Processing of Collected Statistical Data Using the Big Data Processing Method

The developed system allows simultaneous processing of hydrological, climatic, operational, and geospatial datasets within a unified distributed environment, significantly improving forecasting accuracy and reliability [4,5].

2. Materials and Methods

The proposed methodology was implemented using Python programming language and Apache Spark distributed computing environment. The framework integrates multi-source datasets collected from hydrological stations, IoT sensors, SCADA systems, GIS databases, and satellite monitoring platforms [6].

The datasets collected include River discharge; Water level; Flow velocity; Temperature; Rainfall intensity; Snow water equivalent; Glacier melting index; Humidity; Wind speed; Reservoir volume; Turbine discharge; Hydropower generation indicators. All datasets were stored within Hadoop Distributed File System (HDFS), enabling parallel distributed processing of large hydrological datasets (Table 1).

Table 1. Big Data parameters and evaluation criteria

№	Parameter	Formula	Criterion
1	Data volume	$V = \sum S_i$	$V \geq 1\,TB$
2	Data velocity	$V_t = N/\Delta t$	$V_t \geq 10^3$
3	Data variety	$K_v = N_{type}/N_{total}$	$K_v \geq 0{,}3$
4	Data quality	$Q = 1 - N_{error}/N_{total}$	$Q \geq 0{,}95$
5	Economic efficiency	$E = B/C$	$E > 1$
6	Variability	$\sigma^2 = \frac{1}{n}\sum(x_i - \mu)^2$	$\sigma^2 \geq 10$

The processing workflow consists of six interconnected stages: Data collection and storage; Primary preprocessing; Statistical analysis; Distributed computation; Visualization; Evaluation and forecasting. The preprocessing stage includes anomaly detection, normalization, filtering, missing-value reconstruction, and distributed cleaning algorithms using PySpark and Pandas libraries. Because hydrological datasets demonstrate heterogeneous statistical behavior, several probability distribution models were applied during preprocessing: Normal distribution; Log-normal distribution; Poisson distribution; Binomial distribution [7].

The normal distribution model was applied for water level, temperature, humidity, and hydraulic head datasets:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

where: μ— mean value; σ— standard deviation.

For anomaly detection and standardization, Z-score normalization was used:

$$z = \frac{x-\mu}{\sigma} \tag{2}$$

Log-normal distribution was applied for river discharge, turbine flow, and hydropower generation datasets because these parameters contain only positive asymmetric values [9]:

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \tag{3}$$

Poisson distribution was applied for rainfall event analysis:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{4}$$

where: λ— expected event intensity; k— observed number of events (Fig. 2).

For reliability evaluation and binary-state monitoring systems, binomial distribution was used:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{5}$$

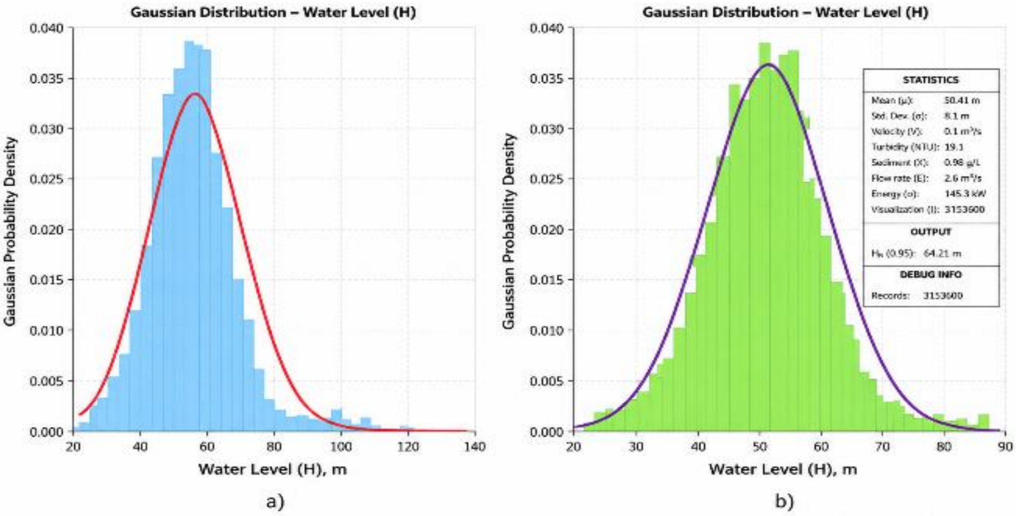

Fig. 2. Results of Primary Processing of Water Level Data Based on Gaussian Distribution Law: a) before processing, b) after processing [8, 9]

3. Result and discussion

The obtained results demonstrate that the proposed GIS-based Big Data processing framework effectively processes heterogeneous hydrological datasets characterized by high variability and distributed spatial structures [Table 2].

Table 2. Integrated Big Data indicators for hydropower monitoring datasets

Parameter	Annual volume	Reliability Q	Economic value E
River discharge	85 MB	0,97	2,8
Water level	420 MB	0,98	2,6
Temperature	210 MB	0,98	1,8
Rainfall	65 MB	0,95	2,1
Glacier melting index	4,5 GB	0,94	2,7
Energy generation	7,8 GB	0,99	3,5
Bathymetry	12 GB	0,93	2,4

The integrated hydrological environment demonstrated all principal Big Data characteristics including high volume, velocity, variety, variability, reliability, and visualization complexity. The total annual data volume exceeded 30,7 GB/year, while the acquisition speed varied from 0,00000038 records/s to 1,0 records/s. The preprocessing results showed substantial improvements in statistical stability and forecasting suitability. Water level datasets initially demonstrated asymmetry with skewness values near 0,85 and kurtosis approximately equal to 1,20. After Gaussian filtering and normalization, skewness decreased to nearly 0,05, indicating close correspondence to normal distribution behavior. River discharge datasets initially contained significant outliers and asymmetric distributions with skewness values approximately equal to 2,1 and kurtosis near 5,8. After logarithmic transformation and filtering, skewness decreased to nearly 0,35, substantially improving forecasting stability [10].

The proposed framework also demonstrated high operational reliability. Data quality coefficients ranged between 0,93 and 0,99, confirming the effectiveness of distributed preprocessing and anomaly detection algorithms. GIS integration enabled simultaneous spatial-temporal analysis of hydrological processes and energy generation dynamics [11].

Distributed PySpark-based processing significantly improved computational efficiency for large-scale hydrological datasets. This capability is especially important for mountainous river basins characterized by rapidly changing hydrological conditions caused by snowmelt, precipitation variability, and climate-related uncertainties .The results obtained confirm that the integration of GIS technologies, Big Data processing environments, and machine learning preprocessing approaches creates an effective scientific basis for long-term forecasting of hydropower resources and sustainable operation of multi-purpose hydropower complexes [12,13].

Conclusions. This study proposed a GIS-based Big Data processing framework for long-term forecasting of mountain river hydropower resources in multi-purpose hydropower complexes. The developed architecture integrates IoT sensors, SCADA systems, GIS technologies, distributed storage environments, and PySpark-based parallel processing methods. The proposed preprocessing framework successfully applied normal, log-normal, Poisson, and binomial statistical distributions for heterogeneous hydrological datasets. The obtained results demonstrated improved data quality, reduced anomaly influence, and increased forecasting stability. The integrated Big Data environment satisfied the principal Big Data criteria including volume, velocity, variety, reliability, variability, and economic efficiency. The developed distributed processing framework significantly improved the accuracy of long-term hydrological forecasting and operational analysis. The obtained results confirm that GIS-integrated Big Data processing technologies provide an effective scientific and practical foundation for sustainable hydropower resource management and intelligent forecasting systems in mountainous regions.

REFERENCES

1. Lu Y, Guo A, Chang J, Wang Y, Wang X, Niu C, Zhang S, Xie Z, Wang J (2025) Enhancing the economic benefits of cascade hydropower stations through integrated utilization of electricity and capacity benefits. Energy 340:139214.
2. Wang Y, Levin T, Kwon J, Baker E (2025) The value of hydropower flexibility for electricity system decarbonization. Energy Reports 13:2711–2721.
3. Тургунбоев ЖК, Ниёзов НН (2025) Искусственный интеллект как инструмент повышения эффективности использования водных ресурсов. In: Цифровая трансформация – шаг в будущее. БГУ, Минск, pp 30–32.
4. Usmonov DS, Kurbonov NN, Ufa RA (2025) Big Data Processing Techniques for Analyzing Collected Statistical Information. In: Бутковские чтения. Томск, pp 140–142.
5. Belina Y, Kebede A, Masinde M (2024) Comparative analysis of HEC-HMS and machine learning models for rainfall-runoff prediction in the upper Baro watershed, Ethiopia. Hydrology Research 55:873–889.
6. Rakhmonov IU, Abduraxmonov IA, Qurbonov NN, Erejepov KD (2026) Design and implementation of an integrated SCADA, IoT, and GIS-based system for real-time monitoring and long-term forecasting in hydropower complexes. Science and Education in Karakalpakstan 145–148.
7. Nguyen T, Nguyen H-T, Nguyen-Hoang T-A (2025) Data quality management in big data: Strategies, tools, and educational implications. Journal of Parallel and Distributed Computing 200:105067.
8. Song C, Cui J, Cui Y, Zhang S, Wu C, Qin X, Wu Q, Chi S, Yang M, Liu J, Chen R, Zhang H (2025) Integrated STL-DBSCAN algorithm for online hydrological and water quality monitoring data cleaning. Environmental Modelling & Software 183:106262.
9. Wang J, Li X, Wu R, Mu X, Baiyinbaoligao, Wei J, Gao J, Yin D, Tao X, Xu K (2025) A runoff prediction approach based on machine learning, ensemble forecasting and error correction: A case study of source area of Yellow River. Journal of Hydrology 658:133190.
10. Hao R, Yan H (2025) Towards Interpreting Machine-Learning Models for Multi-Step Ahead Daily Streamflow Forecasting. Hydrological Processes 39:e70163.
11. Rakhmonov IU, Abdurakhmonov IA (2025) GIS-BASED multi-objective modeling and long-term forecasting of mountain river hydropower potential. Qo‘qon, pp 51–54.
12. Allaev KR, Rakhmonov IU, Mamarasulova TS, Najimatdinov RK (2025) GIS-supported forecasting of hydropower potential for sustainable energy planning in mountainous regions (2020–2030). Universum: технические науки 30–33.
13. Allaev K, Nazirova H, Badalov A (2025) Forecasting of electricity losses using the polynomial regression method. Tashkent, Uzbekistan, p 030008