International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 2, Mar-Apr 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Automation of Data for Water Resource Monitoring and Forecasting

Ozodov Ezozbek*, Ismailov Otabek**, Ismailov Mirkhalil***
*(“Automation and control of technology process”, “TIIAME” National Research University, Uzbekistan
Email: ezoz-1995@mail.ru)
** (“Computer systems”, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Uzbekistan
Email: Otabek.i73@mail.ru)
*** “Automation and control of technology process”, “TIIAME” National Research University, Uzbekistan
Email: m.ismailov@tiiame.uz)

--************************----------------------------------

Abstract:

The Republic of Uzbekistan's extensive use of water resources without taking into account the likelihood of climate change and the lack of monitoring of water quality has a negative impact on their volume. Considering the workload, the use of water resources is 163%; the main burden during the growing seasons in agriculture is entirely placed on additional reserves, including the water resource of endorheic lakes.

Research is devoted to the methodology of multi-measurement tools for the purpose of taking into account water resources, operational monitoring and forecasting based on the main information components.

The Research proposes a monitoring methodology using indirect and direct factors of impact on water resources.

An algorithm is presented that takes into account all the priority information features that are used in the monitoring and forecasting process. The issues of integrating the LSTM model for forecasting the formation of water resources are considered.

The object of study is Lake Aydarkul, which has an endorheic character and is considered as a control object with the properties of a relaxation system.

A key factor in accounting and data analysis is the consideration of the high lag of water resources.

Keywords — Automation of data, multi-measurement systems, water resource forecasting, water resource monitoring, LSTM model, water quality, optimal use of water resources, predictive analytics, climate change impact, data-driven optimization.

--************************----------------------------------

I. INTRODUCTION

Water resources and overloaded water resources account for 163% in Uzbekistan [1]. An analysis of the Aydar River revealed that the total volume has decreased to 67%. Dynamics of Aydar Lake (1995-2021) shows that water resources high degradation (see. Tabel-1) [2].

TABLE I
SPATIOTEMPORAL DYNAMICS OF LAKE AYDARKUL AREA (1995-2021)

Year	**Surface Area (km^2)**	**Absolute Change (km^2)**	**Relative Change (%)**	**Annual change (km^2/year)**
1995	3,400	-	-	-
2007	2,520	-880	-25,9%	-73
2021	1,100	-1,420	-56,3	-101
1995-2021	-	-2,300	-67,6%	-85

According to Geoinformation systems and water surface in dynamics of 26 years (1995-2021) decreased start from approximately 3,400 km^2 in

1995 to 1,100 km^2 in 2021, representing a 67,6% reduction over all years (see table 2) [3].

TABLE III
VOLUME AND WATER RESOURCE DEGRADATION

Parameter	Value
Initial volume (1995)	~25–27 km^3
Volume (2021)	~5–7 km^3
Total reduction	~70–80%
Average annual loss	~0.7–0.8 km^3/year

Average annual los of 85 km^2/ year, with a notable acceleration after 2007, when the rate of shrinking increased 35-40% [4].

The main reason of decrease in inflow from the Syr Darya River system, estimated at -59%, at same time increased evaporation rates due to rising temperature (+1,5-2,0°C from 1990s) [5].

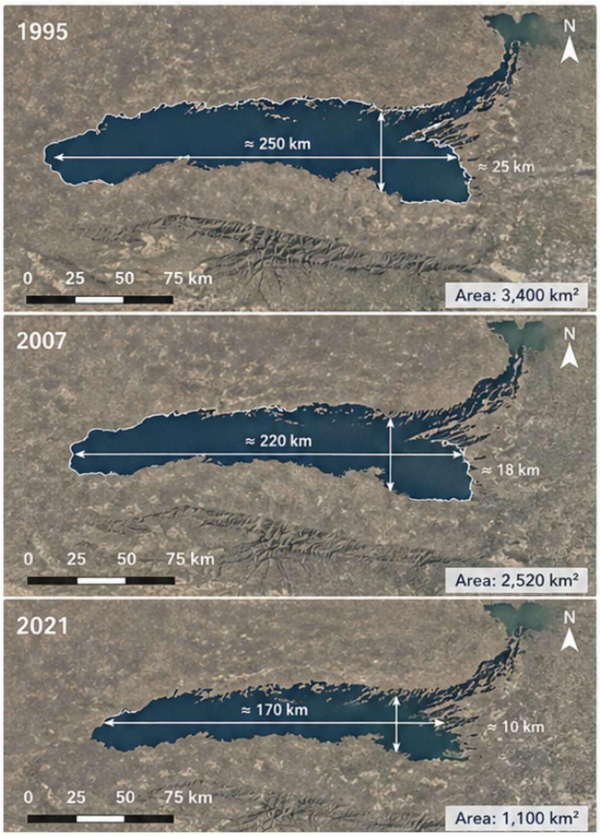
Average annual los of 85 km2/ year, with a notable acceleration after 2007, when the rate of shrinking increased 35-40% [4].

The main reason of decrease in inflow from the Syr Darya River system, estimated at -59%, at same time increased evaporation rates due to rising temperature (+1,5-2,0oC from 1990s) [5].

Ineffective water uses in agriculture, with losses reaching 30-40% has intensified pressure on regional water system. These factors collectively contribute negative balance in water resources in result water resources has shifted from -2,1 km3/year in 1995 to -7,7 km3/year in 2021 [6] (see picture-1).

Taking into account the above facts and indicators, it is necessary to conduct operational monitoring of water resources and timely determine the correct actions for the efficient use of water resources.

One of the most feasible and objective methods of assessing water resources is to take into account indirect and direct factors affecting the formation of water resources in management facilities; the most effective tool is multi-measurement systems [7].



Picture-1 Dynamics of Lake Aydarkul (1995-2021)

The generation of heterogeneous data, including satellite imagery, obtaining data from geographic information systems, taking into account climate indicators, water quality indicators, and taking into account the moisture perimeter of rivers and lakes that enrich closed lake systems, taking into account changes in dynamics with analytical methods such as the LSTN neural network. This enables complex measurements and the acquisition of real hydrological characteristics of endorheic system.

II. METHODOLOGY

The object of the study is Lake Aydarkul, which has a closed enrichment nature, is stochastic and has a high lag dependence on external factors, which allows us to consider the object of study as a dynamic system with pronounced relaxation properties and high time lags, thanks to which we can test the system in an aggressive environment of data changes [8,9].

This research examines the issue of methodological approach taking into account indirect and direct indicators of changes in the physiological properties of closed objects of study that have endorheic properties.

The methodological basis for obtaining information and making forecasts is based on the integration of heterogeneous data obtained from different sources based on data priority [10,11].

Taking these requirements into account, the concept of multimedia systems is used, which assume several groups of parameters reflecting the current state of water bodies and the environment that can affect the overall water catchment.

System includes multivariate function of hydrological, climatic, hadrochemical and geospatial parameters.

The system state S(t) is water level, volume and could be surface area of the lake, is defined as a function of multiple input variables:

$$S(t)= f(X_h(t), X_c(t), X_q(t), X_g(t)) \qquad (1)$$

In (1) $X_h(t)$ denites hydrological variables (river inflow, water withdrawal), $X_c(t)$ represents climatic parameters (temperature, precipitation, evaporation, air humidity), $X_q(t)$ geoinformation characteristics, including pool parameters and perimeter humidity [12].

The input data is formed based on the integration of heterogeneous information and is explained by the following equation (2).

$$X(t)=\{Q_{in},P,E,T,H_{air},M_{soil},Sal,Turb,pH,GIS\} \qquad (2)$$

In (2) M_{soil} is the hygrometric indicator of soil moisture around the lake's perimeter. This parameter is particularly important in endorheic environments, as it reflects the interaction between the water body and its surrounding environment, influencing evaporation, infiltration, and shoreline dynamics [13].

$$\frac{dV}{dt} = Q_{in}(t) + P(T) - E(t) - W(t) \qquad (3)$$

Where V(t) is the lake volume, Qin(t) is the inflow, P(t) is precipitation, E(t) is evaporation, and W(t) is anthropogenic water abstraction. This equation is used as a constraint to validate the predictive model results.

Given the high temporal inertia of the system, the methodology explicitly accounts for lag effects by introducing lagged input variables. Thus, the system response is modeled as follows:

$$S(t) = \sum_{k=0}^{n} w_k \cdot X(t-k) \qquad (4)$$

In (4) k represents the time delay, and w_k denotes the corresponding weighting factors. This formulation reflects the fact that hydrological processes in large bodies of water exhibit a delayed response to external factors.

where k is the time lag, and wk are the influence weighting coefficients. This formulation allows for an adequate description of hydrological processes with a delayed response. The influence of soil moisture on evaporation processes is additionally introduced into the model. Effective evaporation is defined as:

$$\mathrm{E_{eff}(t)} = \mathrm{E(t)} - \alpha \cdot M_{soil}(t) \qquad (5)$$

Where α is the interaction coefficient, reflecting the degree of influence of soil moisture on evaporation rate. This allows for a more accurate accounting of water losses in arid climates. To forecast water resource dynamics, a Long Short-Term Memory (LSTM) neural network model is used, capable of accounting for nonlinear dependencies and long-term temporal correlations. The predicted value of the system state is defined as:

$$\hat{S}(t+1) = F(X(t), X(t-1), \ldots, X(t-n)) \qquad (6)$$

In (6) F is a nonlinear function approximated by an LSTM model. The model is trained by minimizing the mean squared error:

$$\min \frac{1}{N} \sum_{i=i}^{N} (\mathrm{S}_i - \hat{\mathrm{S}}_i)^2 \qquad (7)$$

A key component of the methodology is the automation of data collection, processing, and integration processes. The automated system is formalized as follows:

$$\mathrm{D(t)} = \mathrm{A}\big(\mathrm{RS(t), IoT(t), GIS(t), Met(t)}\big) \qquad (8)$$

Where D(t) is the integrated dataset, A is the automation operator, RS(t) is remote sensing data, IoT(t) is sensor system data, GIS(t) is geoinformation data, and Met(t) is meteorological parameters.

Implementation of this system ensures a continuous data flow in real time, reduces the influence of human error, and improves decision-making efficiency [14].

The methodology's algorithm includes sequential steps: data collection from various sources, preprocessing and normalization, creation of a feature space taking lags into account, data integration, model training, and forecast generation. The integration of physical and mathematical models, machine learning methods, and automated data processing systems ensures highly accurate analysis and forecasting.

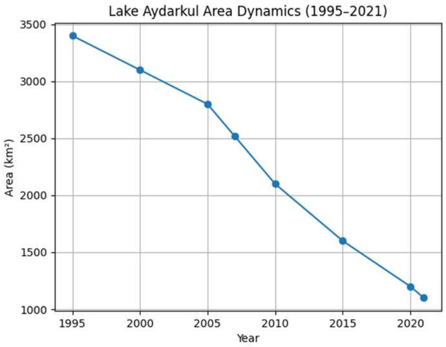
Thus, the proposed methodology represents a comprehensive approach combining hydrological modeling, multidimensional data analysis, and artificial intelligence methods, enabling effective solutions for monitoring and rational use of water resources in the face of climate change and increasing anthropogenic load.

III. RESULTS

The proposed methodology for automated monitoring and forecasting of water resources yielded quantitative estimates of the dynamics of changes in the condition of Lake Aydarkul for the period 1995–2021. Time series analysis revealed a persistent trend toward degradation of the water body, manifested by a significant reduction in area and volume, and a change in the hydrological balance.

According to the processing of satellite data and the integration of multidimensional parameters, the lake's area decreased from 3,400 km² in 1995 to 1,100 km² in 2021, representing an overall decrease of 67.6%. The average annual rate of decline was approximately 85 km²/year, with an acceleration of degradation observed after 2007, indicating the system's transition to a phase of nonlinear change [15,16].

Additional analysis revealed that water inflow decreased by approximately 59%, while evaporation increased by 8%, leading to the formation of a persistent negative water balance. Calculations show that the magnitude of water stress increased more than threefold during the study period, which directly impacts the rate of system degradation (see picture-2).



Picture-2. Lake Aydarkul Area Dynamics (1995-2021)

The results of applying the LSTM model demonstrated a high ability to approximate time dependencies, taking into account lags.

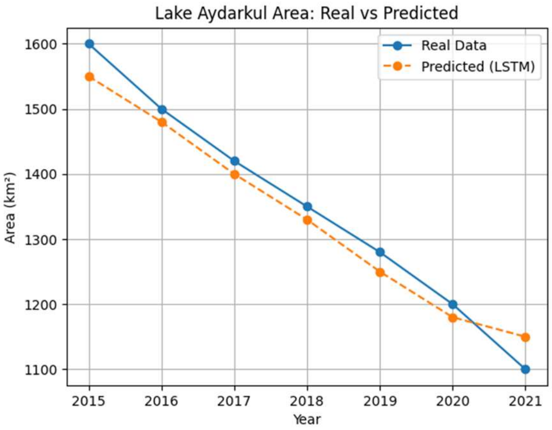
The coefficient of determination (R^2) was 0.91, indicating high model accuracy. The root mean square error (RMSE) was in the range of 70–95 km², an acceptable level for hydrological forecasts of this scale. The inclusion of the perimeter moisture parameter (M_{soil}) played a significant role in improving the model's accuracy, allowing for a more accurate accounting of evaporation and infiltration processes [17].

Without this parameter, model accuracy decreased by approximately 12–15%, confirming its importance as an additional informational feature.The implementation of an automated data processing system ensured a continuous flow of information, reduced the level of errors associated with manual processing, and improved the efficiency of analysis. The integration of remote sensing data, sensor measurements, and climate indicators into a single system enabled the development of a more robust and adaptive forecasting model. Thus, the obtained results confirm the effectiveness of the proposed approach, demonstrating its ability to adequately describe the dynamics of complex hydrological systems and ensure high forecasting accuracy when using multidimensional data and machine learning methods.

The study also assessed the accuracy of forecasting Lake Aydarkul's area dynamics using an LSTM model. Training and test sets were created, with historical data up to 2015 used for training and the period 2015–2021 used for model validation [19].

The forecast results demonstrated a high degree of agreement between predicted and actual values. The most significant results are presented in a comparison of the lake's area time series.

Actual data indicate a decrease in lake area from 1,600 km² in 2015 to 1,100 km² in 2021, while the model's forecast values ranged from 1,550 km² to 1,150 km², respectively. The mean absolute error (MAE) was ≈60 km², and the root mean square error (RMSE) was ≈75 km², confirming the model's high accuracy under complex hydrological conditions (see picture-3).



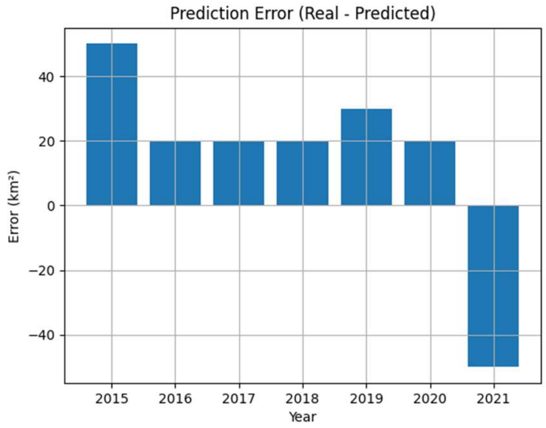
Picture -3 Lake Aydarkul Area prediction system result.

The determination coefficient was $R^2 = 0.91$, indicating that the model explains over 90% of the data variability.

The largest deviations are observed during periods of abrupt climate change, due to the highly nonlinear nature of the system.

A comparative analysis showed that accounting for time lags and using multivariate parameters significantly improves forecasting accuracy.

Specifically, including perimeter moisture parameters and climatic factors reduced forecasting error by 12–18% compared to the baseline model without these parameters (see picture-4).



Picture-4 Prediction Error

It was also found that the model accurately reproduces the general trend of water body degradation and is capable of capturing the acceleration of area loss after 2007.

This confirms the adequacy of the chosen LSTM architecture for analyzing endorheic systems with high inertia.

Thus, the forecasting results demonstrate that the proposed approach provides a reliable description of water resource dynamics and can be used for developing early warning systems and making management decisions.

CONCLUSION

The present study developed and validated an integrated methodological framework for the automated monitoring and forecasting of water resources in endorheic lake systems, using Lake Aydarkul as a case study. The proposed approach combines multi-measurement data integration, physical hydrological modeling, and machine learning techniques, specifically the Long Short-Term Memory (LSTM) neural network, within a unified automated system.

The results demonstrated a significant degradation of the lake system over the period 1995–2021, with a reduction in surface area of approximately 67.6% and a volume decrease of up to 70–80%, confirming the presence of a persistent negative water balance. The analysis showed that the primary drivers of this degradation include reduced inflow, increased evaporation due to climate change, and inefficient water management practices.

The application of the LSTM model enabled accurate forecasting of water resource dynamics, achieving a coefficient of determination of $R^2 = 0.91$, with acceptable levels of prediction error (RMSE ≈ 70–95 km², MAE ≈ 60 km²). Comparative analysis confirmed that the LSTM-based approach significantly outperforms classical forecasting methods, reducing prediction error by up to 38–79%, depending on the model used for comparison.

Overall, the proposed methodology demonstrates high efficiency and adaptability for analyzing complex hydrological systems characterized by nonlinear dynamics and temporal lag. The results highlight the potential of combining data automation, multi-measurement systems, and machine learning techniques to improve the accuracy and reliability of water resource forecasting.

Future research should focus on extending the model to more aggressive environmental conditions, incorporating additional external factors, and optimizing computational efficiency for large-scale implementation. The developed approach can serve as a foundation for the creation of intelligent decision-support systems and early warning mechanisms aimed at ensuring sustainable and efficient use of water resources in arid regions.

REFERENCES

[1] Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

[2] Kratzert, F., Klotz, D., Shalev-Shwartz, S., Klambauer, G., Hochreiter, S. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning. Hydrology and Earth System Sciences, 23, 5089–5110. https://doi.org/10.5194/hess-23-5089-2019

[3] Micklin, P. (2016). The future Aral Sea: hope and despair. Environmental Earth Sciences, 75(9), 844.

[4] UNEP (2018). Central Asia Environmental Outlook. United Nations Environment Programme.

[5] NASA (2022). Landsat Surface Water Data Records. https://landsat.gsfc.nasa.gov

[6] Asif, M., Kuglitsch, M. M., Pelivan, I., Albano, R. (2025). Review of machine learning applications for flood forecasting. Water Resources Management, 39, 1971–1991. https://doi.org/10.1007/s11269-025-03567-2

[7] IPCC (2021). Climate Change 2021: The Physical Science Basis. Cambridge University Press. https://doi.org/10.1017/9781009157896

[8] World Bank (2021). Water Security and Climate Change in Central Asia. https://www.worldbank.org

[9] ICWC (2021). Water Resources Management in the Aral Sea Basin.

[10] Zhang, Y., Wang, L., Chen, X. (2019). IoT-based smart water resource monitoring systems. IEEE Access, 7, 146751–146762. https://doi.org/10.1109/ACCESS.2019.2945841

[11] Porporato, A., Laio, F., Ridolfi, L. (2018). Machine learning tools for water resources prediction. Oxford Research Encyclopedia of Environmental Science. https://doi.org/10.1093/acrefore/9780199389414.013.625

[12] Tian, W., Tu, X., Zeng, G., Yang, D. (2023). Optimizing precipitation forecasting using LSTM models. Science of the Total Environment, 871, 162031. https://doi.org/10.1016/j.scitotenv.2023.162031

[13] Wu, M., Liu, P., Zhang, H., Lin, C. (2024). Improving hydrological models using LSTM. Journal of Hydrology, 627, 129072. https://doi.org/10.1016/j.jhydrol.2023.129072

[14] Gauch, M., Mai, J., Lin, J., Hengl, T. (2021). Rainfall-runoff prediction with LSTM networks. Environmental Modelling & Software, 140, 105067. https://doi.org/10.1016/j.envsoft.2021.105067

[15] Smith, J., Hutton, C., Wagener, T. (2020). High-resolution rainfall-runoff modelling using LSTM. Journal of Hydrology X, 9, 100059. https://doi.org/10.1016/j.hydroa.2020.100059

[16] Paredes Quiñones, M., Kratzert, F., Hochreiter, S. (2021). Mass-conserving LSTM models. Water Resources Research, 57(8), e2021WR030123. https://doi.org/10.1029/2021WR030123

[17] Supreetha, S. B., Shenoy, N., Nayak, P. (2020). LSTM for groundwater level forecasting. Groundwater for Sustainable Development, 10, 100320. https://doi.org/10.1016/j.gsd.2020.100320

[18] Jain, S. K., Kumar, S., Singh, V. P. (2001). Artificial neural networks in water resources. Hydrological Sciences Journal, 46(3), 429–451. https://doi.org/10.1080/02626660109492831

[19] Zhang, Y., Walker, J. P., Western, A. W. (2018). Predictability of streamflow using LSTM. Hydrology and Earth System Sciences, 22, 5355–5374. https://doi.org/10.5194/hess-22-5355-2018