

Artificial Intelligence-Assisted Water Quality Model: Long-Term Follow-up Data

Yeldos A. Altay

*Department of Robotics and Technical Means
of Automation
Satbayev University
Almaty, Kazakhstan*

Lashin Bazarbay

*Department of Robotics and Technical Means
of Automation
Satbayev University
Almaty, Kazakhstan
lashyn_7754@mail.ru*

Abstract— This article presents the results of developing a model for assessing water quality using the artificial intelligence method. The presented model is based on linear regression, which, when evaluated, revealed a statistically significant interdependence between the combined water quality indicators. It was found that among the measured parameters, the most influential predictor of acidity, conductivity, turbidity and redox potential is water temperature. The discovered relationship between the measured indicators is mainly associated with the influence of temperature on the physical and chemical processes that occur when the temperature of river water increases and decreases.

Keywords— *water quality control, artificial intelligence, linear regression, temperature, acidity, turbidity, conductivity.*

I. INTRODUCTION

Water is one of the main vital elements in living systems and a resource for terrestrial and aquatic ecosystems. The influence of various factors, namely anthropogenic and environmental, can significantly change the biological, physical and chemical properties of water, reducing its quality. Taking this into account, water quality control is currently carried out briefly in laboratory conditions and on an ongoing basis using specialized measuring systems, including Internet of Things (IoT) devices for remote transmission of measured data. Among the listed types of water quality monitoring, it is long-term monitoring that can reveal a number of significant changes that cannot be detected in a timely manner with short-term monitoring. In this regard, long-term monitoring of water quality in real time is an important scientific problem that requires an integrated approach to solution.

Based on the principles of forming an integrated approach, in this work, to develop a model for assessing water quality, long-term monitoring is carried out using modern sensors that allow measuring temperature (t), turbidity (TUF), acidity (pH), electrical conductivity (C) and redox potential (ORP, e). It is important to note that the joint measurement of five indicators for water quality control has not previously received proper development, which is confirmed by scientific results published in [1]–[15]. This can mainly be attributed to the fact that acidity, turbidity and conductivity sensors are used to jointly assess water quality, while other researchers combine only temperature, without an ORP sensor. A joint assessment of water parameters together with ORP will determine the oxygen content and decomposition of pollutants [1], [7]. Thus, the assessment of the ORP indicator together with the indicators listed above, as well as their combination for water quality control, is important and necessary, as they will allow for a

comprehensive assessment of the indicators and identify significant associations between them when developing a model for assessing water quality.

The development of a model for assessing water quality while monitoring indicators is carried out using artificial intelligence methods. It is known that the use of artificial intelligence methods makes it possible to increase the reliability and develop a predictive model of the controlled object when assessing water quality. According to published data [2], [8], among artificial intelligence methods for developing a predictive model for assessing water quality, linear and nonlinear regression, as well as their combination with the empirical mode decomposition method, are used. This hybrid implementation makes it possible to develop a mathematical model that characterizes the relationship between the physical and chemical parameters of water. However, it should be noted that the empirical mode decomposition method and its various modifications when decomposing measured data can cause a mode mixing effect due to “edge effects” [16], [17], which can significantly complicate the process of developing an assessment model based on the artificial intelligence.

This article proposes a model for assessing water quality based on the artificial intelligence method - linear regression. This method is effective and highly sensitive to the analyzed data, which does not require special specific training of samples from various control objects [18], [19]. Thus, the purpose of the research in this article is to develop a mathematical regression model that predicts water quality and the association of aggregate indicators.

II. COMBINED MEASUREMENT DATA COLLECTION

In this section of the article, for the combined assessment of indicators, the hardware of the measuring system has been developed, which implements a combined measurement of temperature, turbidity, acidity, electrical conductivity and redox potential of water on the ATmega2560 microcontroller platform and provides wireless transmission of measured data to a special Blynk mobile application via a cloud service and IoT devices. The result of the development of a measuring system for collecting and remotely transmitting data on combined water indicators is presented in Fig. 1.

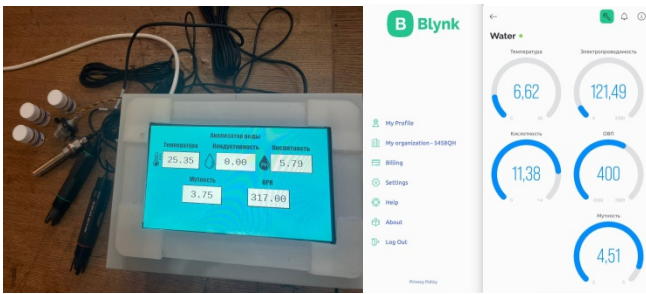


Fig. 1. System for collecting and transmitting combined measurement data

It is important to note that the measuring system presented in Fig. 1, unlike known analogues, allows for combined measurement of five indicators of water from various control objects. When testing this system, the collection of measurement data was carried out on the basis of a three-month observation of water indicators in the waters of the Almaty River in the city of Almaty. It is important to note that the first observation was carried out from September 1 to September 30, the second observation was carried out from October 1 to October 31, and the third observation was carried out from November 1 to 30, respectively. Each of the observations consists of more than $n = 10000$ measurements samples and have differences in measurement values. The data set of combined measurements generated in this way contributes to the development of a mathematical model and the identification of changes in the quality of the water from which they were obtained.

When developing a regression model, the most sensitive and prognostic indicator of the process of changing the properties of substances and materials is the average value of the measurement results. In this regard, in the work under review, to develop a linear regression model for assessing water quality, the choice of the average value of the results consisting of more than $n = 10000$ measurements is justified. The result of the model development is presented in the next section of this article.

III. REGRESSION MODEL FOR WATER QUALITY ASSESSMENT

The model for assessing water quality in regression analysis, depending on the results of measuring combined indicators, is divided into linear, nonlinear and logistic analysis [8]. The listed analysis methods make it possible to develop a model for assessing water quality with a certain degree of reliability, which can be characterized quantitatively. In accordance with the results of the study [18], [19], the reliability of the model in this analysis can be characterized through the coefficient of determination R^2 , the significance level p -value and the relationship between the observed and predicted values through the correlation coefficient r .

The calculation results demonstrated that the average value of three-month indicators linearly increases and decreases depending on the quality of river water. The obtained result allows us to justify the choice of linear regression for the purpose of constructing regression dependence and assessing the association of joint indicators using the least squares method. In Fig. 2–5 shows the results of regression (analytical) dependence of water quality indicators are presented.

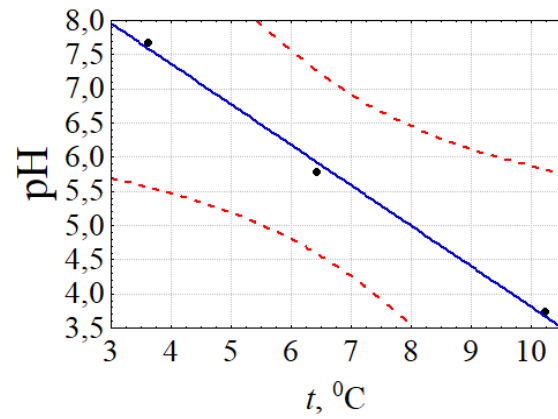


Fig. 2. Analytical dependence temperature and acidity

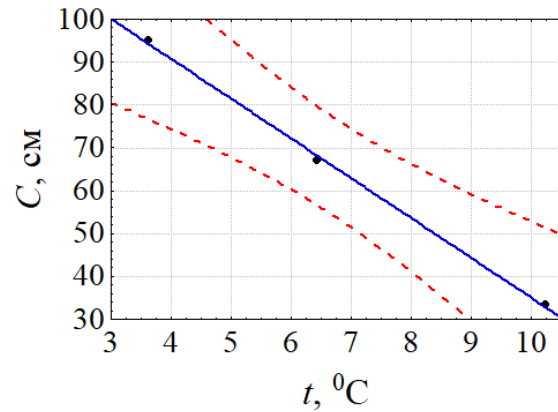


Fig. 3. Analytical dependence temperature and conductivity

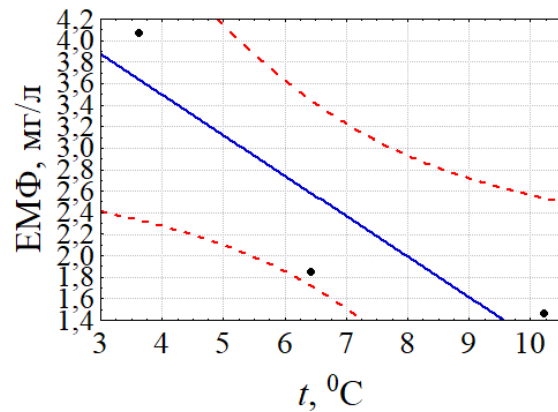


Fig. 4. Analytical dependence temperature and turbidity

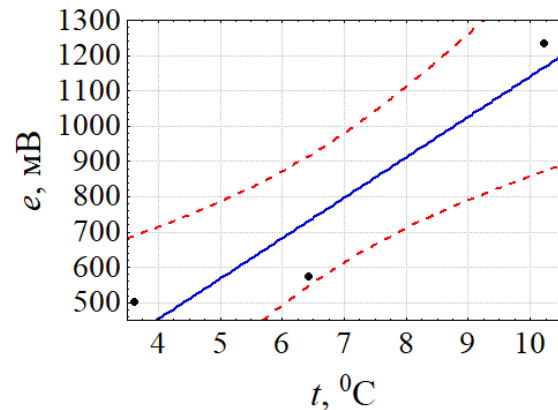


Fig. 5. Analytical dependence temperature and ORP

From the presented dependence results in Fig. 2–5 it follows that due to an increase in the temperature of the analyzed river water samples, the values of acidity, conductivity and turbidity of water decrease, therefore, the value of the oxidative reduction potential of water increases depending on the observation time.

The regression model characterizing the influence of the predictor – temperature on the indicators of acidity, conductivity, turbidity and redox potential of the analyzed river water samples is described by a linear regression equation and the coefficients of determination R^2 and correlation r , as well as the significance p -level of the model are calculated as:

$pH = -0,5922 \cdot t + 9,732$ при $R^2 = 0,995$; $r = -0,997$; $p = 0,0414$ (Fig. 2);

$C = -9,283 \cdot t + 127,885$ при $R^2 = 0,998$; $r = -0,999$; $p = 0,022$ (Fig. 3);

$EM\Phi = -0,376 \cdot t + 5,003$ при $R^2 = 0,791$; $r = -0,889$; $p = 0,302$ (Fig. 4);

$e = 114,571 \cdot t + 4,678$ при $R^2 = 0,884$; $r = 0,940$; $p = 0,220$ (Fig. 5);

The obtained result of the regression analysis allows us to predict that with an increase in the temperature of the analyzed river water samples by 1 degree, we should expect a decrease in water acidity by 99.58%, conductivity by 99.87%, turbidity by 79.10%, as well as an increase in redox potential by 88.49%, respectively. The established association is quite close to a linear change and is resistant to the influence of river water temperature. It is important to note that the share of other influencing factors when constructing a regression model, namely acidity accounts for 0.42%, conductivity – 0.13%, turbidity 20.9%, ORP – 11.51%.

The coefficient determination value R^2 of the developed model explains the observed variance of the predictor (the indicator in the focus of this study) when monitoring water quality indicators. The resulting model coefficients with a p -value less than that for $p = 0.05$ are therefore considered statistically significant for the confidence interval with a reliability of $P = 0.95$, respectively. That is, related combined indicators can be significant predictors of controlled processes occurring in water with a probability of confidence interval reliability. To date, accumulated experimental material on assessing water quality confirms the possibility of using selected indicators to carry out predictive analytics of the state of water samples [8].

IV. CONCLUSION

This article presents the results of developing a mathematical regression model that predicts water quality and the association of aggregate indicators. The results showed that as river water temperature increased, there was a decreasing trend in the relationship between acidity, conductivity and turbidity, with a correlation coefficient ranging from $r = -0.889$ to $r = -0.999$. It should also be noted that as the water temperature increases, the value of the ORP indicator increases with the correlation coefficient $r = 0.940$, respectively.

The results obtained are partly consistent with the results of foreign studies, in which remotely sensed indicators proposed in this work competed as the most influential predictors for a water quality assessment model.

REFERENCES

- [1] Tariq M. O. An Open Source Water Quality Measurement System for Remote Areas. *Engineering Proceedings*. 2021. № 1. P. 50.
- [2] Zhao Y. Retrieval of water quality parameters based on near-surface remote sensing and machine learning algorithm. *Remote Sensing*. 2022. № 21. P. 5305.
- [3] Wood R. Implementation of the WHO core components of an infection prevention and control programme in two sub-saharan African acute health-care facilities: a mixed methods study. *Antimicrobial Resistance and Infection Control*. 2024. № 1. P. 4.
- [4] Mathur R. Modeling of two-stage anaerobic onsite wastewater sanitation system to predict effluent soluble chemical oxygen demand through machine learning. *Scientific Reports*. 2024. № 1. P. 1835.
- [5] Wang Z. Research on water environmental indicators prediction method based on EEMD decomposition with CNN-BiLSTM. *Scientific Reports*. 2024. № 1. P. 1676.
- [6] Shah M. I., Javed M. F., Abunama T. Proposed formulation of surface water quality and modelling using gene expression, machine learning, and regression techniques. *Environmental Science and Pollution Research*. 2021. № 28. P. 13202-13220.
- [7] de Camargo E.T. Low-cost water quality sensors for IoT: A systematic review. *Sensors*. 2023. № 9. P. 4424.
- [8] Essamlali I., Nhaila H., El Khaili M. Advances in machine learning and IoT for water quality monitoring: A comprehensive review. *Heliyon*. 2024.
- [9] Fadel A. A., Shujaa M. I. Water Quality Monitoring System Based on IOT Platform. *IOP Conference Series: Materials Science and Engineering*. 2020. № 3. P. 032054.
- [10] Bogdan R. Low-cost internet-of-things water-quality monitoring system for rural areas. *Sensors*. 2023. № 8. P. 3919.
- [11] Lakshmikantha V. IoT based smart water quality monitoring system. *Global Transitions Proceedings*. 2021. № 2. P. 181-186.
- [12] Geetha S., Gouthami S. Internet of things enabled real time water quality monitoring system. *Smart Water*. 2016. № 2. P. 1-19.
- [13] Shah M. I., Javed M. F., Abunama T. Proposed formulation of surface water quality and modelling using gene expression, machine learning, and regression techniques. *Environmental Science and Pollution Research*. 2021. № 28. P. 13202-13220.
- [14] Chowdury M.S. IoT based real-time river water quality monitoring system. *Procedia computer science*. 2019. № 155. P. 161-168.
- [15] Sabari M. Water quality monitoring system based on IoT. 2020 IEEE 5th International Conference on Devices, Circuits and Systems. 2020. P. 279-282.
- [16] Altay Y.A. Cascade Notch Filter with a Unity Feedback and Improved Transient Response. *V International Conference on Control in Technical Systems*. 2023. P. 217-220.
- [17] Y.A. Altay, A. S. Kremlev, "Signal-to-Noise Ratio and Mean Square Error Improving Algorithms Based on Newton Filters for Measurement ECG Data Processing," in *Proc. EIConRus*, 2021, paper 21.09.04, pp. 1590-1595.
- [18] Fedorov A.V., Altay Y.A., Stepanova K.A., Kuzivanov D.O. The effect of signal-to-noise-ratio value on the error in measuring acoustic emission parameters: statistical assessment. *Scientific and technical journal of information technologies, mechanics and optics*, 2022, V. 22, pp. 1205-1215.
- [19] Y.A. Altay, A.S. Kremlev, "On the use of the statistical methods for biomedical signals and data processing" in *Proc. EIConRus*, 2019, paper 19.04.03, pp. 1129-1134.