UDC 004.8:631.6.03
DOI: 10.15587/2706-5448.2025.323730

**Altay Yeldos,
Lashyn Bazarbay,
Kassymbek Ozhikenov,
Zhandos Dosbayev,
Ulantay Nakan,
Zhuldyz Kalpeyeva**

CORRELATION AND REGRESSION ANALYSIS IN ASSESSING THE RELATIONSHIP BETWEEN WATER INDICATORS: A BRIEF DESCRIPTION OF LONG-TERM MEASUREMENT DATA FROM BIOSENSORS

The object of the study is the method for assessing the relationship between the results of long-term observations of water parameters obtained by the method of combined measurements by a biosensor system. The biosensor system is designed for the combined measurement of five water parameters based on physical value sensors. In the paper, the problem under consideration quite fully levels out a significant limitation of the known solutions designed for the simultaneous measurement of three or four water parameters. Existing approaches in their structure combine less than five biosensors-sensors, which significantly limits the simultaneous measurement of five water parameters.

One of the main and principal results of the paper is the development of a statistical model for assessing the relationship between the combined measurements of five water parameters. It was revealed that among the five measured parameters, the most influential predictor for acidity, conductivity, turbidity and oxidation-reduction potential is water temperature. The established significant and non-random relationship between the parameters is mainly associated with the effect of temperature on the physical processes occurring with an increase and decrease in water temperature depending on the observation time. These estimates demonstrate a higher, statistically significant relationship between the measurement information data. This is achieved by implementing the method of aggregate measurement of water parameters: temperature, acidity, turbidity, conductivity, oxidation-reduction potential.

The efficiency of the statistical model is confirmed by calculating the correlation coefficient based on the Pearson method and the coefficients of determination and reliability of the model. The regression model can be used in practice in developing new or improving known measuring systems and control devices to increase the reliability and effectiveness of water quality control.

***Keywords:** water indicator, measurement system, multi-sensors system, correlation analysis, testing, internet of things, biosensors.*

Received: 30.11.2024
Received in revised form: 04.02.2025
Accepted: 16.02.2025
Published: 26.02.2025

How to cite
Yeldos, A., Bazarbay, L., Ozhikenov, K., Dosbayev, Z., Nakan, U., Kalpeyeva, Z. (2025). Correlation and regression analysis in assessing the relationship between water indicators: a brief description of long-term measurement data from biosensors. Technology Audit and Production Reserves, 1 (2 (81)), 49–53. https://doi.org/10.15587/2706-5448.2025.323730

1. Introduction

Water is one of the main vital elements in living systems and a resource for terrestrial and aquatic ecosystems, and its quality is affected by various anthropogenic and environmental factors that alter its physical, biological and chemical properties. The three types of water properties listed above are evaluated by measuring indicators based on a class of analytical control methods. These methods include chemical, physicochemical, and physical analyses. Biological and chemical properties are assessed for the presence of bacteriological contamination, identification of chemical compounds, including macro- and micronutrients, while the assessment of physicochemical properties is aimed specifically at identifying processes that may occur specifically in the aquatic environment [1–14].

Physicochemical analysis methods at the present stage of research are implemented by measuring a number of quantitative indicators based on biosensors that have the ability to fully reflect water quality in order to identify processes occurring specifically in the aquatic environment. Therefore, research aimed at studying the physicochemical properties of water based on an intelligent multisensory – biosensors system that increases the information content of cumulative measurement data and the effectiveness of monitoring is a modern scientific and technical task.

Monitoring of water parameters based on an intelligent biosensor system is a critical moment for timely control of water pollution, the importance of which lies in its innovative approach to monitoring using specialized biosensors. It is important to note that traditionally used methods of monitoring water indicators require significant resources, time and specialized analytical measuring equipment, which limits their availability, especially in remote regions and places.

In reviews [8, 9], it is noted that one of the main and important tasks of water quality control when measuring indicators is the study and search for predictors that reflect physical processes. Taking into account the noted, in numerous studies [6–14] the main attention is paid to the importance of forming a set of measurements based on the development of a measuring system in order to search for predictors and assess the relationship of indicators characterizing the physical processes of water.

 ISSN-L 2664-9969; E-ISSN 2706-5448

In work [7], an experimental setup was developed for analytical control of physical processes of water, combining in its structure temperature, turbidity and acidity sensors on a modern element base, an ATMega2560 microcontroller and an Internet of Things module. The implementation of the setup made it possible to state the fact of the influence of temperature and turbidity on water quality.

In works [6, 10–13], a measuring system for water quality control was developed, containing sensors of physical quantities, characterized in that sensors for determining the oxidation-reduction potential of water, a conductivity sensor and a microcontroller in combination with a module for wireless data transmission to a cloud service for subsequent analysis using artificial intelligence methods are additionally installed.

The works [8, 9] show that a measuring system combining in its structure sensors measuring parameters based on potentiometry and optical turbidometry, combined with each other, allows obtaining water quality control results during analytical control.

The conducted analytical review of the literature revealed [1–13] that in recent years, work in this direction has been carried out on the development of a multisensor system in combination of three or four sensors – biosensors. Researchers involved in the development of a control system based on three sensors mainly select the parameters of acidity, turbidity and temperature, while another group of researchers expands these parameters by adding a sensor – biosensor and measuring the oxidation-reduction potential or temperature. One of the main features of the listed multisensor systems is that it is impossible to simultaneously measure the oxidation-reduction potential and temperature, on which the quality of water depends. For example, in works [6–14] it is noted that the measured water temperature together with the above indicators can be considered as an energy measure of water molecules, affecting the physicochemical properties of water, thereby its quality.

The development of a model for assessing water quality by 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 of the observed object and develop a predictive model for water quality assessing. According to scientific data [2, 8], a prediction model for water quality assessment is developed using artificial intelligence techniques such as linear and nonlinear regression in conjunction with the empirical mode decomposition approach. The hybrid implementation of the considered methods makes it possible to develop a mathematical model describing the relationship between the physical and chemical parameters of water. In addition, the method of empirical mode decomposition of measured indicators (DMI) – empirical modal decomposition of measurement results and various modifications of this method causes the Gibbs effect [15]. As a result, the process of developing an evaluation model based on the artificial intelligence method becomes more complicated.

In the paper under review, an artificial intelligence method is proposed – a model of water quality assessment based on linear regression. This method is characterized by efficiency and high sensitivity to the analyzed data that does not require specific machine learning of samples from different observation objects [16, 17]. Thus, *the aim of this research* is to develop a regression model that assumes the relationship between water quality indicators. The peculiarity of the reviewed article from the previous publication [16, 17] is that temperature factor effects on water quality indicators in autumn and spring months are illustrated together based on the linear regression method.

2. Materials and Methods

This section of the paper presents the development of the hardware part of a measuring system designed for a comprehensive analysis of water quality. The material is a set of aggregate measurements of water parameters, measured and formed at the output of the developed multi-sensor control system equipped with five physical quantity sensors. The set of aggregate measurements of these water parameters is formed on the basis of the developed multi-sensor control system in [16, 17]. The multi-sensor system performs measurements of temperature, formazine turbidity, acidity, conductivity, and redox potential. The development is based on the ATmega2560 microcontroller, which ensures data collection and processing. The obtained data is transmitted wirelessly via IoT devices to a cloud service. Subsequently, the information is sent to a dedicated mobile application Blynk for convenient monitoring and analysis.

The described microcontroller features make it possible to collect combined water measurement data and apply it to remote data transmission using special wireless modules. This is evidenced by recent scientific publications [1, 2, 6, 7, 9, 12–14].

Measurements of combined water indicators were carried out in the autumn and spring months in the water area of the Almaly River in Almaty (Republic of Kazakhstan) based on the developed measurement system in scientific articles published in previous years [16, 17]. The design of the water indicators control in autumn and spring months are illustrated in Fig. 1.

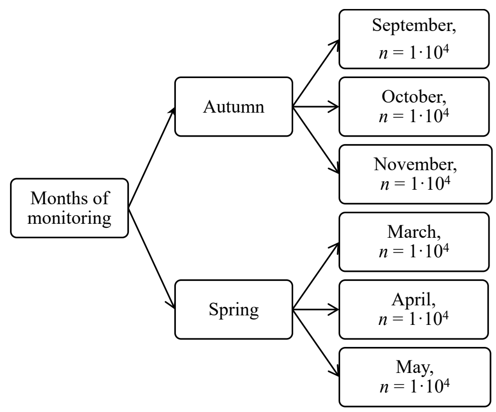

Fig. 1. The design of water indicators in autumn and spring

In the paper, the first monitoring of the autumn months was conducted from 1 to 30 September, the second monitoring was conducted from 1 to 31 October and the third monitoring was conducted from 1 to 30 November, as shown in Fig. 1. The first observation in the spring months was from 1 to 31 March, the second from 1 to 30 April and the third from 1 to 31 May. The number of measurements of each of the five water indicators we mentioned above exceeds $n=10000$ and consists of heterogeneous differences. The set of combined measurements thus formed is based on the development of a mathematical model.

It is known that the most sensitive and predictable indicator of the process of variation of the properties of substances and materials in the development of a regression model is the average value of the measurement result.

In this regard, this paper justifies the calculation and estimation of the mean value of results consisting of more than $n=10{,}000$ measurements for the development of a linear regression model for water quality assessment. The result of the regression model development is presented in the next section of the paper.

The correlations between the measurements were assessed using Pearson's method and were performed as follows:

$$\widehat{r}_{ij} = \frac{\sum_{q=1}^{N}(\widehat{n}_i^{\langle i\rangle}(q)-\overline{n^{\langle i\rangle}}(q))(\widehat{n}_i^{\langle j\rangle}(q)-\overline{n^{\langle j\rangle}}(q))}{\sqrt{\sum_{q=1}^{N}(\widehat{n}_i^{\langle i\rangle}(q)-\overline{n^{\langle i\rangle}}(q))^2\sum_{q=1}^{N}(\widehat{n}_i^{\langle j\rangle}(q)-\overline{n^{\langle j\rangle}}(q))^2}}, \quad (1)$$

where $\hat{n}_i^{\langle i \rangle}$ – current value of i-th indicator; $\overline{n^{\langle i \rangle}}$ – average value of i-th indicator; $\hat{n}_i^{\langle j \rangle}$ – current value of j-th indicator; $\overline{n^{\langle j \rangle}}$ – average value of j-th indicator; n – number of measurement samples.

Interpretation of the strength and direction of the relationship was carried out using a Chaddock specialized evaluation scale, where the relationship is considered noticeable at values of Pearson correlation coefficient $r=0.3–0.7$, at $r=0.7–0.9$ high and at $r=0.9–1$ very high [15]. The difference of indicators is considered statistically significant at the level of $p=0.05$ for the confidence interval with reliability $P=0.95$, respectively in the assessment of measurement results.

3. Results and Discussion

In regression analysis, the water quality assessment model is subdivided into linear, non-linear and logistic analysis depending on the results of measuring the combined indicators [8]. The listed methods of analysis allow developing a water quality assessment model with a certain degree of reliability that can be quantitatively characterized. According to the results of the study [17], the reliability of the model in this analysis is described by the relationship between the observed and predicted values through the coefficient of determination R^2, significance level p-value and correlation coefficient r.

The results of computations demonstrated that the average values of indicators for autumn and spring months linearly rise and fall depending on water quality. The obtained result allows to justify the selection of linear regression in order to build regression dependence and assess the relationship of indicators using the method of least squares. Fig. 2–5 below show the results of regression dependence of water quality indicators on the base of scattergram.

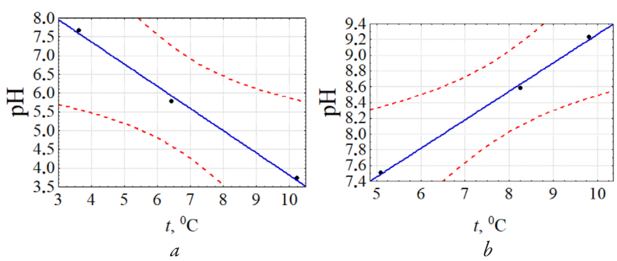

Fig. 2. Dependence of temperature effect on acidity: *a* – by autumn months; *b* – by spring months

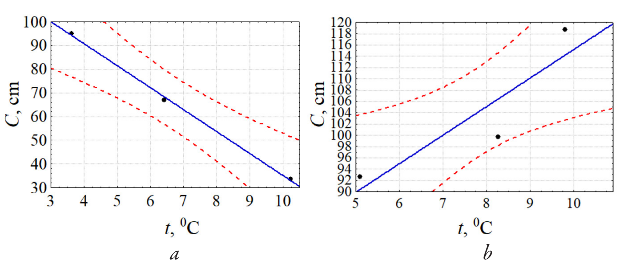

Fig. 3. Dependence of temperature effect on conductivity: *a* – by autumn months; *b* – by spring months

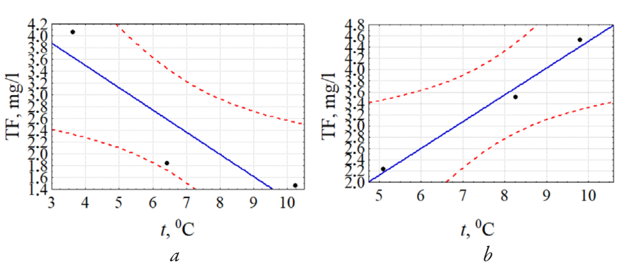

Fig. 4. Dependence of temperature effect on turbidity: *a* – for autumn months; *b* – for spring months

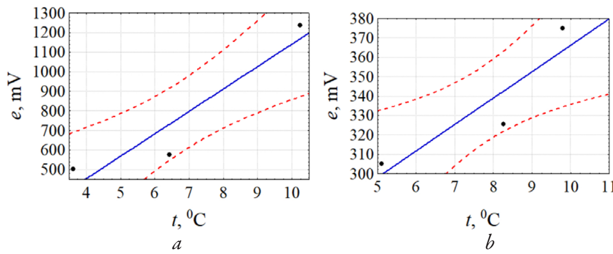

Fig. 5. Dependence describing the influence of redox potential with temperature: *a* – by autumn months; *b* – by spring months

The proposed results of the dependence in Fig. 2–5 indicate that in the analyzed water samples due to the increase in temperature the values of acidity, conductivity and formazine turbidity of water decrease, so depending on the observation time the value of redox processes of water increases. The model of analyzed water samples describing the influence of the predictor – temperature on acidity pH, conductivity C, formazine turbidity FT and redox processes e is written by linear regression equation and determination coefficients R^2 and correlation r, as well as significance level p-model as follows:

– Fig. 2, *a*:

$$pH_{autumn}=-0.5922t+9.732;\ R^2=0.995;\ r=-0.997;\ p=0.0414;$$

– Fig. 3, *a*:

$$C_{autumn}=-9.283t+127.885;\ R^2=0.998;\ r=-0.999;\ p=0.022;$$

– Fig. 4, *a*:

$$FT_{autumn}=-0.376t+5.003;\ R^2=0.791;\ r=-0.889;\ p=0.302;$$

– Fig. 5, *a*:

$$e_{autumn}=114.571t+4.678;\ R^2=0.884;\ r\,0.940;\ p=0.220;$$

– Fig. 2, *b*:

$$pH_{spring}=0.3622t+5.6456;\ R^2=0.9968;\ r=0.9984;\ p=0.0360;$$

– Fig. 3, *b*:

$$C_{spring}=5.064t+64.5691;\ R^2=0.8063;\ r=0.8980;\ p=0.2901;$$

– Fig. 4, *b*:

$$FT_{spring}=0.4754t+0.2491;\ R^2=0.9831;\ r=0.9915;\ p=0.0831;$$

– Fig. 5, *b*:

$$e_{spring}=13.6082t+230.1194;\ R^2=0.8253;\ r=0.9085;\ p=0.2745,$$

where R^2 is the coefficient of determination; r is the correlation coefficient; p is the significance level of the model.

The regression analysis obtained for the autumn month suggests that if the temperature of the analyzed water samples increases by 1 degree Celsius, it is possible to expect a decrease in water acidity by 99.58 %, conductivity by 99.87 %, formazine turbidity by 79.10 % and an increase in redox index by 88.49 %. Correspondingly, spring indicators will increase as follows: it is possible to expect 99.68 % decrease in acidity, 80.63 % decrease in conductivity, 98.31 % decrease in formazine turbidity and 82.53 % increase in redox. Very close to the observed linear change in the joint and resistant to water temperature. In the construction of regression model of water quality determination by autumn and spring months, the proportion of influencing extraneous factors – for acidity – 0.42 % and 0.32 %, for conductivity – 0.13 % and 19.37 %, for formazine turbidity level – 20.9 % and 1.69 % and for redox – 11.51 % and 17.47 %.

It should be noted that the explanation of the correlation coefficient calculated using the Karl Pearson method is characterized by a direct and high correlation between the measured indicators using the Cheddock scale. For these indicators, the established correlation at exceeding $r > 0.7$ explains the difference in the values of water indicators and the existence of a non-random fundamental relationship between the indicators.

The considered dependence of the indicators on temperature, presented in Fig. 2–4, is opposite to each other based on the observation period, namely during the autumn months. It was in the autumn months, unlike in the spring months, that the amount of precipitation tended to increase, which consequently led to an increase in the acidity of the aquatic environment. This, in turn, contributed to the emergence of a negative correlation between the indicators of acidity, turbidity, and conductivity. However, the presence of a negative correlation does not imply the absence of a relationship but merely characterizes its direction.

It is also important to note that no increase in the amount of precipitation affecting the aquatic environment indicators was observed during the spring months, which is also consistent with meteorological data [16, 17]. Therefore, by comparing these dependencies, it is possible to assert that the inverse correlation of autumn water measurements is influenced by the amount of precipitation, which can be interpreted as the impact of environmental factors on the effectiveness of this monitoring. Nevertheless, despite this, water temperature remains one of the important indicators involved in the physicochemical processes in aquatic environment monitoring objects [1, 2, 7].

It is important to note that the value of the coefficient of determination R^2 of the proposed regression model explains the observed variance of the predictor (the indicator in the focus of this study) in monitoring water quality indicators. The obtained model coefficients have a p-value less than $p = 0.05$ and hence are considered statistically significant for confidence interval with probability $p = 0.95$. At the same time, the water quality indicators in a binding combination with confidence interval probability can be recognized as important predictors of the observed processes occurring in water. At present, the accumulated experimental material on water quality assessment confirms the possibility of using the selected indicators for predictive analyses of water samples [8].

The qualitative indicators selected in the paper to determine water quality are labelled as predictive indicators and in the next research work, based on these indicators and in combination with additional predictive indicators, the antibiotic indicators in water will be determined. Obviously, in future research work, the five selected indicators in this paper and the two additional indicators that will be selected for future research work will be implemented together with an IoT module based on a special multi-sensor information measurement system.

The features obtained results is that, with the combined measurement of the indicators, a statistically significant correlation link was obtained using a multi-sensor control system, confirming the association between the selected indicators. Previously, a significant association of indicators reflecting the physical processes of water with a combined measurement was insufficiently studied. Moreover, the results obtained under consideration provide new ideas about the physical processes of water with a combined measurement of five indicators, and also provide opportunities to predict the properties of substances at the stage of analytical control of liquid media.

The results discussed above provide new insights into the physical processes involved in the combined measurement of five water parameters. Moreover, these results can be widely applied in practice to sort data by predictor in the task of finding an association between five parameters. The prospects for further development are the creation or special training of measured data for decision making in order to display recommendation information.

4. Conclusions

The paper presents the results of the development of a mathematical model of regression dependence assuming the relationship between water quality indicators. The results obtained showed that as water temperature increases, the correlation coefficient from $r = -0.889$ to $r = -0.998$ tends to decrease the relationship between acidity, conductivity and turbidity indices by formazine. Also, as the water temperature increases during autumn and spring months, respectively, the value of redox indicator increases according to the correlation coefficient interval $r_{autumn} = 0.940$ and $r_{spring} = 0.908$.

Financing

This research has been/was/is funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (BR24993051 Development of an intelligent city system based on IoT and data analysis).

Conflict of interest

The authors declare that they have no conflict of interest in relation to this study, including financial, personal, authorship or other, which could affect the study and its results presented in this article.

Data availability

The manuscript has no associated data.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies in creating the presented work.

References

1. Tariq, M. O., Siddiq, A., Irshad, H., Aman, M., Khan, M. S. (2021). An Open Source Water Quality Measurement System for Remote Areas. *The 1st International Conference on Energy, Power and Environment, 50.* https://doi.org/10.3390/engproc2021012050
2. Zhao, Y., Yu, T., Hu, B., Zhang, Z., Liu, Y., Liu, X. et al. (2022). Retrieval of Water Quality Parameters Based on Near-Surface Remote Sensing and Machine Learning Algorithm. *Remote Sensing, 14 (21),* 5305. https://doi.org/10.3390/rs14215305
3. Wood, R., Tembele, W., Hema, A., Somé, A., Kinganda-Lusamaki, E., Basilubo, C. et al. (2024). 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 & Infection Control, 13 (1).* https://doi.org/10.1186/s13756-023-01358-1
4. Mathur, R., Sharma, M. K., Loganathan, K., Abbas, M., Hussain, S., Kataria, G. et al. (2024). Modeling of two-stage anaerobic onsite wastewater sanitation system to predict effluent soluble chemical oxygen demand through machine learning. *Scientific Reports, 14 (1).* https://doi.org/10.1038/s41598-023-50805-x
5. Wang, Z., Duan, L., Shuai, D., Qiu, T. (2024). Research on water environmental indicators prediction method based on EEMD decomposition with CNN-BiLSTM. *Scientific Reports, 14 (1).* https://doi.org/10.1038/s41598-024-51936-5
6. Shah, M. I., Javed, M. F., Abunama, T. (2020). Proposed formulation of surface water quality and modelling using gene expression, machine learning, and regression techniques. *Environmental Science and Pollution Research, 28 (11),* 13202–13220. https://doi.org/10.1007/s11356-020-11490-9
7. de Camargo, E. T., Spanhol, F. A., Slongo, J. S., da Silva, M. V. R., Pazinato, J., de Lima Lobo, A. V. et al. (2023). Low-Cost Water Quality Sensors for IoT: A Systematic Review. *Sensors, 23 (9),* 4424. https://doi.org/10.3390/s23094424

8. Essamlali, I., Nhaila, H., El Khaili, M. (2024). Advances in machine learning and IoT for water quality monitoring: A comprehensive review. *Heliyon, 10 (6),* e27920. https://doi.org/10.1016/j.heliyon.2024.e27920
9. Abbas Fadel, A., Ibrahim Shujaa, M. (2020). Water Quality Monitoring System Based on IOT Platform. *IOP Conference Series: Materials Science and Engineering, 928 (3),* 032054. https://doi.org/10.1088/1757-899x/928/3/032054
10. Bogdan, R., Paliuc, C., Crisan-Vida, M., Nimara, S., Barmayoun, D. (2023). Low-Cost Internet-of-Things Water-Quality Monitoring System for Rural Areas. *Sensors, 23 (8),* 3919. https://doi.org/10.3390/s23083919
11. Lakshmikantha, V., Hiriyannagowda, A., Manjunath, A., Patted, A., Basavaiah, J., Anthony, A. A. (2021). IoT based smart water quality monitoring system. *Global Transitions Proceedings, 2 (2),* 181–186. https://doi.org/10.1016/j.gltp.2021.08.062
12. Geetha, S., Gouthami, S. (2016). Internet of things enabled real time water quality monitoring system. *Smart Water, 2 (1).* https://doi.org/10.1186/s40713-017-0005-y
13. Chowdury, M. S. U., Emran, T. B., Ghosh, S., Pathak, A., Alam, Mohd, M., Absar, N. et al. (2019). IoT Based Real-time River Water Quality Monitoring System. *Procedia Computer Science, 155,* 161–168. https://doi.org/10.1016/j.procs.2019.08.025
14. Sabari, M., Aswinth, P., Karthik, T., Bharath Kumar, C. (2020). Water Quality Monitoring System Based on IoT. *2020 5th International Conference on Devices, Circuits and Systems (ICDCS),* 279–282. https://doi.org/10.1109/icdcs48716.2020.243598
15. Altay, Y. A., Lyamin, A. V., Kelemseiit, N. E., Skakov, D. M. (2023). Cascade Notch Filter with a Unity Feedback and Improved Transient Response. *2023 V International Conference on Control in Technical Systems (CTS),* 217–220. https://doi.org/10.1109/cts59431.2023.10288775
16. Altay, Y. A., Bazarbay, L., Uskenbayeva, R. K., Fedorov, A. V., Kalpeyeva, Z. B. (2024). Water Indicator Measurement and Data Processing From Internet of Things Modules. *2024 International Conference on Electrical Engineering and Photonics (EExPolytech),* 253–256. https://doi.org/10.1109/eexpolytech62224.2024.10755886
17. Altay, Y. A., Bazarbay, L. (2024). Artificial Intelligence-Assisted Water Quality Model: Long-Term Follow-Up Data. *2024 V International Conference on Neural Networks and Neurotechnologies (NeuroNT),* 1–3. https://doi.org/10.1109/neuront62606.2024.10585363

Altay Yeldos, *PhD, Candidate of Sciences, Senior Lecturer, Department of Robotics and Technical Means of Automation, Satbayev University, Almaty, Republic of Kazakhstan, ORCID: https://orcid.org/0000-0002-3736-0291*

✉***Lashin Bazarbay,*** *Senior Lecturer, Department of Robotics and Technical Means of Automation, Satbayev University, Almaty, Republic of Kazakhstan, e-mail: l.bazarbay@satbayev.university, ORCID: https://orcid.org/0000-0002-8694-8622*

Kassymbek Ozhikenov, *Candidate of Sciences, Professor, Head of Department of Robotics and Technical Means of Automation, Satbayev University, Almaty, Republic of Kazakhstan, ORCID: https://orcid.org/0000-0003-2026-5295*

Zhandos Dosbayev, *PhD, Senior Lecturer, Department of Electronics, Telecommunications and Space Technologies, Satbayev University, Almaty, Republic of Kazakhstan, ORCID: https://orcid.org/0000-0003-1673-4036*

Ulantay Nakan, *PhD, Associate Professor, Department of Chemical and Biochemical Engineering, Satbayev University, Almaty, Republic of Kazakhstan, ORCID: https://orcid.org/0000-0002-3758-3962*

Zhuldyz Kalpeyeva, *PhD, Associate Professor, Director of Institute of Automation and Information Technologies, Satbayev University, Almaty, Republic of Kazakhstan, ORCID: https://orcid.org/0000-0002-4970-3095*

✉*Corresponding author*