Modeling the content of chloroform in drinking water infiltration intake by periods of the annual cycle: low-water and permanent watercourse

Maria A. Malkova*, *Nikita* D. Minchenkov, *Olga* G. Kantor, and *Evgeniy* A. Kantor

Ufa State Petroleum Technological University, Ufa, Russia

Abstract. The results of a study to find a relationship between the chloroform content in drinking water and the parameters characterizing water quality over an eighteen-year period of the infiltration water intake operation are presented. To increase the reliability of the model, the annual cycle is divided into two periods: an extended flood, including four months (April - June) and a period of low water and permanent watercourse, including eight months (August - March). Three time series were formed from the initial data: from the monthly average values of the parameters for the entire observation period; as a result of averaging the parameter values corresponding to each year; average values characterizing the low-water period and permanent watercourse for the entire observation period. It was found that the period of low water and permanent watercourse can be described by regression equations characterized by a smaller value of the average approximation error and a large value of the correlation coefficient. It was revealed that April introduces a significant stochasticity in the annual period. It is shown that the results obtained can be used to assess the value of the chloroform content in drinking water.

1 Introduction

One of the goals of sustainable development is to ensure the rational use of water resources. Drinking water resources of good quality and in sufficient quantity are essential for all aspects of life and sustainable development. Water resources are integrated into all forms of development (e.g. health care, food security, etc.), sustainable economic growth in agriculture, industry and energy production, as well as maintaining healthy ecosystems. The implementation of high-quality drinking water to the consumer is a paramount task.

One of the main reasons that determine the relatively low level of agreement between the calculated and experimentally determined values of the chloroform (CF) content in drinking water, possibly, is the high proportion (usually above 62%) of the time series random component of both CF and water quality parameters. which are used as an argument in regression equations [1, 2, 16].

In this regard, we made an attempt to model the CF content in drinking water of infiltration water intake by periods of the annual cycle. The first paper is devoted to the

---

* Corresponding author: physicspoems@yandex.ru

analysis of forecasting the CF content in the extended flood period, including four months - April, May, June and July [16]. The identification of the flood period is determined by the fact that it is the most stochastic in the annual cycle in terms of changes in organoleptic and generalized indicators of water quality parameters used in modeling (turbidity (T), chromaticity (C), oxidizability (O)).

The period of low water and permanent watercourse, including eight months of the annual cycle (from August to March, inclusive), is characterized by stable values of water discharge in comparison with floods.

This article presents the results of statistical modeling. The main task of modeling is to assess the change in the drinking water quality when conditions change in the catchment area and to assess the likelihood of events occurrence that pose a threat to the water supply system (in terms of drinking water quality) [3-7].

2 Materials and methods

Three time series were formed as the initial data describing the low-water period and permanent watercourse: the first one includes the monthly average values of the CF content (for the period 1997-2014) and water quality parameters of the water source (144 for each value); the second is obtained as a result of averaging the parameter values corresponding to each year (18 values) and the third, used for comparison, obtained from the average values (8 values). A detailed description of the time series formation is given earlier [16].

Part 1 presents the facts underlying the justification for separating the extended flood period from the annual cycle. [16]. These include the stochasticity of water discharge in the water source and the presence of inflection points on the curves of the parameter's changes (CF, T, C, O) in the period from April to July [16]. During the low-water period and permanent watercourse, including eight months (from August to March, inclusive), the average water discharge in the water source is about 72% of the average annual and is more than two times less than during the extended flood period (Figure 1).

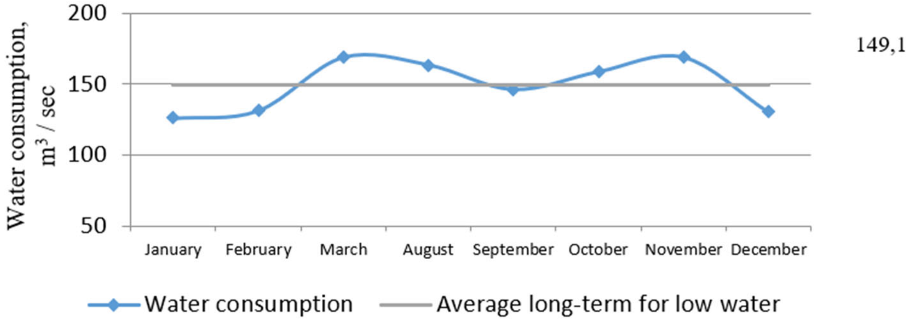

Fig. 1. Water discharge in a water source during low-water period and permanent watercourse, m^3/s.

During the extended flood period, the water discharge reaches 407 m^3 / s, during the indicated 8-month period it does not exceed 170 m^3 / s. Comparison of the average annual water discharge shows that in the annual cycle, the period of extended flood and the period of low water and permanent watercourse, the water discharge is 206.3, 320.6, and 149.1 m^3 / s, respectively. It should be noted that, in comparison with the low-water period (126 m^3 / s), during the period under consideration, the highest flow rate is 168.7 m^3 / s. In addition, the flow rate of water during this period does not have pronounced drops, and its change occurs smoothly. A slight increase in discharge (by 10-15 m^3 / s) occurs in March (due to the spring flood) and in October - November. This is due to the operation of the hydroelectric power plant in connection with the need to increase the supply of electricity due to the onset of the heating season.

Thus, it is advisable to distinguish from the annual cycle an eight-month period (from August to March, inclusive), which covers the period of low water and permanent watercourse.

To determine the influence of the analyzed period on the indicators of the regression equations, the following operation was carried out. The values of CF content and water quality parameters related to the months of the extended flood period were sequentially removed from the annual cycle, and the results obtained were compared with each other and describing the annual period (Table 1).

Correlation-regression analysis was used to process two time series of water quality parameters changes. The first represents the monthly average values of the parameters for the entire observation period (144 values), the second is obtained as a result of averaging the values of the parameters corresponding to each year (18 values). For comparison, the data obtained as average values characterizing the period of low water and permanent watercourse for the entire observation period are presented (8 values).

Table 1. Accepted designations for chloroform content, turbidity, chromaticity and oxidizability of water in equations 1 – 24.

Excluded month	Annual cycle	April	April, May	April, May, June	April, May, June, July
Time series (144 values)	Values characterizing the period of low water and permanent watercourse for the entire observation period				
	$[CF]_1$, $[T]_1$, $[C]_1$, $[O]_1$	$[CF]_2$, $[T]_2$, $[C]_2$, $[O]_2$	$[CF]_3$, $[T]_3$, $[C]_3$, $[O]_3$	$[CF]_4$, $[T]_4$, $[C]_4$, $[O]_4$	$[CF]_5$, $[T]_5$, $[C]_5$, $[O]_5$
Time series (18 values)	Average values of the parameters corresponding to each year for the entire observation period				
	$[CF]_{a1}$, $[T]_{a1}$, $[C]_{a1}$, $[O]_{a1}$	$[CF]_{a2}$, $[T]_{a2}$, $[C]_{a2}$, $[O]_{a2}$	$[CF]_{a3}$, $[T]_{a3}$, $[C]_{a3}$, $[O]_{a3}$	$[CF]_{a4}$, $[T]_{a4}$, $[C]_{a4}$, $[O]_{a4}$	$[CF]_{a5}$, $[T]_{a5}$, $[C]_{a5}$, $[O]_{a5}$
Time series (8 values)	Average parameter values for the entire observation period				
	$[CF]_{ap1}$, $[T]_{ap1}$, $[C]_{ap1}$, $[O]_{ap1}$	$[CF]_{ap2}$, $[T]_{ap2}$, $[C]_{ap2}$, $[O]_{ap2}$	$[CF]_{ap3}$, $[T]_{ap3}$, $[C]_{ap3}$, $[O]_{ap3}$	$[CF]_{ap4}$, $[T]_{ap4}$, $[C]_{ap4}$, $[O]_{ap4}$	$[CF]_{ap5}$, $[T]_{ap5}$, $[C]_{ap5}$, $[O]_{ap5}$

3 Results

We have made an attempt to describe the CF content in drinking water obtained at the infiltration water intake using multivariate correlation-regression analysis. The indicators of the water quality of the source were used as independent variables: turbidity (T), chromaticity (C), oxidizability (O). The initial data used in the calculations represent the results of the quality analytical control of the water source and drinking water over an 18-year period.

Earlier it was indicated that one of the reasons for the high proportion of a random variable in the CF content time series in drinking water can be a shift in the beginning, a change in the duration and intensity of floods in different years [16]. In this regard, it seems appropriate to trace the influence of this period on the change in water quality parameters. For this purpose, a consistent exclusion from the annual cycle of the water quality parameters values (T, C, O) and CF content for April, May, June and July was carried out (Figure. 1a-e).

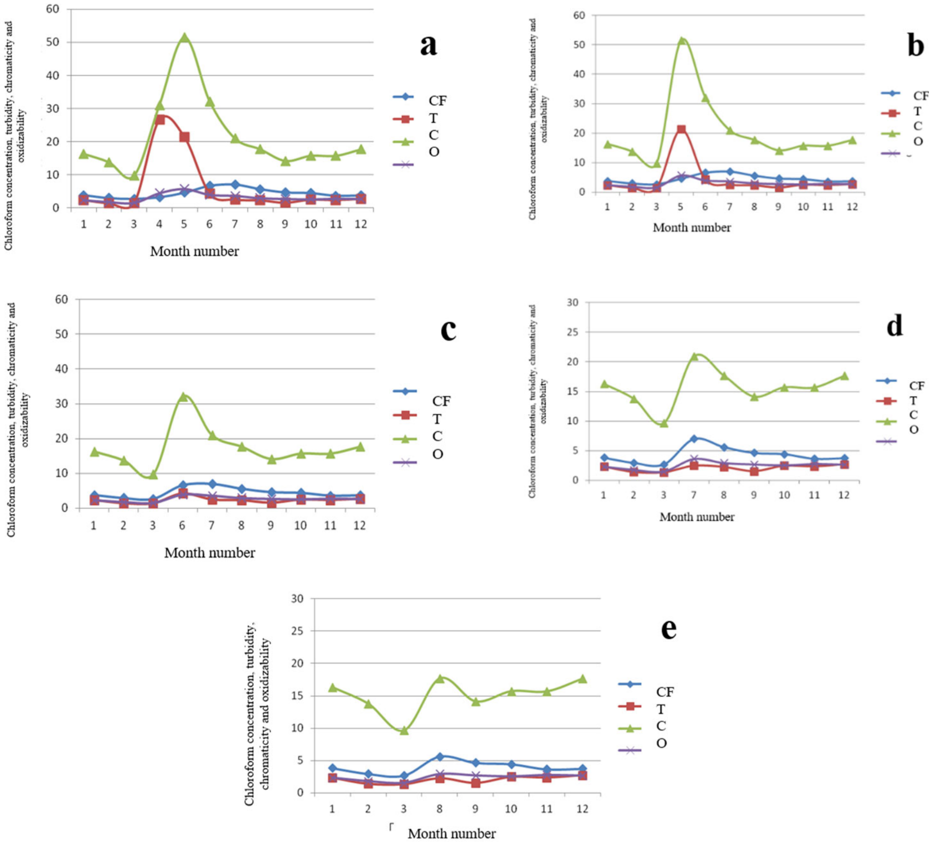

Fig. 2. Average values of chloroform concentration (μg / dm^3), turbidity (mg / dm^3), chromaticity (^{0}) and oxidizability (mg / dm^3) for the entire observation period: a) annual cycle; b) excluding April; c) excluding April and May; d) excluding April, May and June; e) excluding April, May, June and July.

It should be emphasized that comparison of changes in CF content is delayed in comparison with the maximums of C, O, and T by 2 - 3 months (Figure 1a). For CF, the maximum falls on July, for C, O - in May, and for turbidity - in April (Figure 1a). The exclusion of April from the annual cycle leads to the fact that the maximum values of all parameters (C, O, and T) fall on May (Figure 1b), while the maximum value of the CF content is in July, i.e. the delay is 2 months (Figure 1b). If two months are excluded from the annual cycle, the maximum values of the parameters coincide (June), and the CF content is delayed by one month, i.e. falls in July (Figure 1c). The exclusion of three and four months from the annual period (April-June and April-July) predetermines the coincidence of the maximum values of all parameters and CF content (Figure 1d-e). Various changes in parameters revealed during the study can be explained as follows. Before the onset of the flood, the reservoir's water reserves are quite high. To smooth the load, the reservoir operation technology provides for its partial emptying before the snow cover melts and melt water flows into the reservoir [8-11]. This operation is systematic and performed annually [13]. The beginning of the emptying of the reservoir falls on February. At this time, the water temperature is 3 ° C and the change in water quality is primarily determined by the mechanical disruption of bottom and coastal sediments. As a result, T rises significantly, while the growth of C and O values is less pronounced. In the following months, the water temperature rises, as a result of which bioprocesses begin to occur, which lead to the bloom of phytoplankton [14, 15].

Comparison of the average annual values of CF, T, C and O for the period 1997 - 2014 (low water and permanent watercourse) shows that these parameters change less significantly (Figure 2) than during the extended flood period. So, for the time series of average annual CF values, the range of change varies from 1.2 to 7.1 µg / dm^3, turbidity from 1.2 to 3.5 mg / dm^3, for chromaticity - from 6.9 to 24.7 ^{0}, oxidizability - from 1.7 to 3.2 mg / dm^3. While in the extended flood period, these parameters vary from 1.5 to 15 µg / dm^3, from 1.3 to 63.9 mg / dm^3, from 8.5 to 106.9 ^{0}, O 2 to 8, 6 mg / dm^3, respectively (part 1).

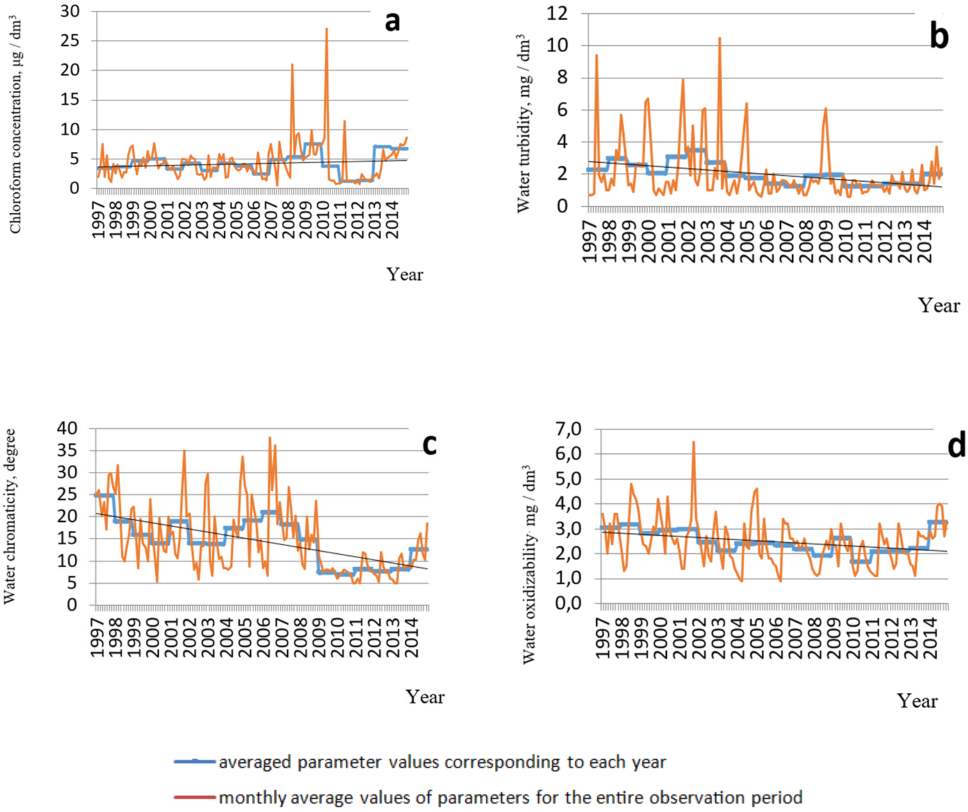

Fig. 3. Average annual values of indicators for the period 1997-2014: a) chloroform, µg / dm^3; b) turbidity, mg / dm^3; c) chromaticity, degrees; d) oxidizability, mg / dm^3.

Analysis of the data representing the mean long-term values shows that equations 1 and 2 are statistically insignificant (Table 2a). In all likelihood, the reason for this is the lack of statistical data, the same reasons lead to the fact that equations 4,5,6 should not be considered reliable (Table 2b). For paired ratios (Table 2c), according to monthly average observational data, equations 7 and 9 are significant, however, they are also characterized by relatively low correlation coefficients (r = 0.59 and 0.36). The relationship between CF and O (Table 2a) is characterized by a high correlation coefficient (r = 0.82), but this result should be considered random.

Table 2. Equations of dependence of monthly average and averaged concentrations of chloroform [CF] on turbidity [T], chromaticity [C], oxidizability [O].

a) according to the average values of the parameters, for the entire observation period (8 values)					b) according to the averaged values of the parameters corresponding to each year (18 values)				
$[CF]_{ap} = 2.45 + 0.72\ [T]_{cr}$				(1)	$[CF]_a = 1.50 + 0.54\ [T]_a$				(4)
A = 17.7	F = 1.20	$R^2 = 0.17$	S = 0.93		A = 40.0	F = 0.79	$R^2 = 0.05$	S = 1.73	
$[CF]_{ap} = 0.38 + 0.24\ [C]_{cr}$				(2)	$[CF]_a = 3.30 - 0.05\ [C]_a$				(5)
A = 14.5	F = 4.41	$R^2 = 0.42$	S = 0.77		A = 55.8	F = 0.38	$R^2 = 0.02$	S = 1.75	
$[CF]_{ap} = 0.17 + 1.55\ [O]_{cr}$				(3)	$[CF]_a = -0.99 + 1.45\ [O]_a$				(6)
A = 9.1	F = 12.62	$R^2 = 0.68$	S = 0.58		A = 40.8	F = 2.83	$R^2 = 0.09$	S = 1.63	
c) according to the monthly average values of the parameters for the entire observation period (144 values)									
$[CF] = 0.75 + 0.92\ [T]$									(7)
A = 49.2	F = 77.87	$R^2 = 0.35$	S = 2.16						
$[CF] = 2.35 + 0.002\ [C]$									(8)
A = 54.4	F = 0.34	$R^2 = 0{,}002$	S = 2.69						
$[CF] = 0.003 + 1.06\ [O]$									(9)
A = 71.4	F = 22.15	$R^2 = 0.13$	S = 2.50						

A significant part of the multiple regression equations (Tables 3, 4, 5), describing different time periods, are characterized by low values of the Fisher criterion (levels 14, 15, 17-19), and therefore, the data summarized in Tables 3 and 4, cannot be considered reliable (statistically significant).

Table 3. Multiple regression equation for the relationship between chloroform content and parameters: turbidity [T], chromaticity [C], oxidizability [O] (8 values).

Excluded month	Equation				N
Annual cycle	$[CF]_{ap1} = -0.22 - 0.19\ [T]_{ap1} - 0.09\ [C]_{ap1} + 2.52\ [O]_{ap1}$				(10)
	A = 10.7	F = 13.03	$R^2 = 0.83$	S = 0.67	
April	$[CF]_{ap2} = -0.60 - 0.36\ [T]_{ap2} + 0.02\ [C]_{ap2} + 2.06\ [O]_{ap2}$				(11)
	A = 8.1	F = 18.05	$R^2 = 0.89$	S = 0.57	
April, May	$[CF]_{ap3} = -0.48 - 1.08\ [T]_{ap3} + 0.11\ [C]_{ap3} + 2.05\ [O]_{ap3}$				(12)
	A = 9.5	F = 19.44	$R^2 = 0.83$	S = 0.55	
April, May, June	$[CF]_{ap4} = -0.48 - 1.08\ [T]_{ap4} + 0.11\ [C]_{ap4} + 2.05\ [O]_{ap4}$				(13)
	A = 10.6	F = 11.91	$R^2 = 0.88$	S = 0.60	
April, May, June, July	$[CF]_{ap5} = -0.07 - 0.82\ [T]_{ap5} + 0.09\ [C]_{ap5} + 1.74\ [O]_{ap5}$				(14)
	A = 9.9	F = 3.93	$R^2 = 0.75$	S = 0.63	

Table 4. Multiple regression equation for the relationship between chloroform content and parameters: turbidity [T], chromaticity [C], oxidizability [O] (18 values).

Excluded month	Equation				N
Annual cycle	$[CF]_{a1} = -0.80 + 0.18\ [T]_{a1} - 0.14\ [C]_{a1} + 2.05\ [O]_{a1}$				(15)
	A = 37.7	F = 1.95	$R^2 = 0.30$	S = 1.59	
April	$[CF]_{a2} = -1.03 + 1.09\ [T]_{a2} - 0.05\ [C]_{a2} + 0.60\ [O]_{a2}$				(16)
	A = 10.5	F = 16.42	$R^2 = 0.78$	S = 0.76	
April, May	$[CF]_{a3} = -1.62 + 0.48\ [T]_{a3} - 0.11\ [C]_{a3} + 1.91\ [O]_{a3}$				(17)
	A = 32.7	F = 1.98	$R^2 = 0.30$	S = 1.59	
April, May, June	$[CF]_{a4} = -1.87 + 0.07\ [T]_{a4} + 0.15\ [C]_{a4} + 2.61\ [O]_{a4}$				(18)
	A = 35.6	F = 2.52	$R^2 = 0.35$	S = 1.52	
April, May, June, July	$[CF]_{a5} = -0.80 + 0.18\ [T]_{a5} - 0.14\ [C]_{a5} + 2.05\ [O]_{a5}$				(19)
	A = 37.8	F = 1.95	$R^2 = 0.30$	S = 1.59	

Table 5. Multiple regression equation for the relationship between chloroform content and parameters: turbidity [T], chromaticity [C], oxidizability [O] (144 values).

Excluded month	Equation				N
Annual cycle	$[CF]_1 = 2.44 - 0.08\ [T]_1 + 0.02\ [C]_1 + 0.67\ [O]_1$				(20)
	A = 42.5	F = 17.05	$R^2 = 0.21$	S = 1.85	
April	$[CF]_2 = 2.35 - 0.09\ [T]_2 + 0.02\ [C]_2 + 0.74\ [O]_2$				(21)

	A = 41.8	F = 16.52	$R^2 = 0.22$	S = 1.87	
April, May	**$[CF]_3 = 1.08 - 0.002\ [T]_3 + 0.03\ [C]_3 + 1.06\ [O]_3$**				(22)
	A = 36.1	F = 34.34	$R^2 = 0.40$	S = 1.68	
April, May, June	**$[CF]_4 = 0.94 - 0.09\ [T]_4 + 0.04\ [C]_4 + 1.09\ [O]_4$**				(23)
	A = 36.8	F = 29.52	$R^2 = 0.39$	S = 1.63	
April, May, June, July	**$[CF]_5 = 1.34 - 0.08\ [T]_5 + 0.04\ [C]_5 + 0.89\ [O]_5$**				(24)
	A = 37.9	F = 19.79	$R^2 = 0.32$	S = 1.45	

The regression equations for the time series of the true values of the parameters have a lower value of the coefficient of determination (0.21 - 0.40) (Table 5) compared to the annual cycle (0.75-0.89) (Table 3). The average approximation error is 36 - 43%. The calculation of the extensive indicators of the equation terms the shows that the free term of the equation and the oxidizability of the water in the source make a greater contribution to the chloroform content (Table 5).

Thus, equations 20-24 (Table 5) can be considered important for the analysis. In general, we can assume that equations 20 and 21 are practically the same, as well as equations 22-24 among themselves.

Comparison with the equation for multiple correlation of CF content with water quality parameters (equation 7, 16) with equation 22-24 shows that the average the error in the second case decreased by 21.5%, while the coefficient of determination increased from 0.12 to 0.39.

In addition, it can be considered that the exclusion from the time series of data related to the onset of the active flood period (April) increases the strength of the relationship between CF content and water quality parameters.

4 Conclusion

Results of determining the possibility of predicting chloroform content using regression analysis depending on organoleptic (turbidity, chromaticity) and generalized (oxidizability) water quality indicators of a water source using the operation of dividing the annual cycle into two periods: extended flood period (April-July) and low-water period and permanent watercourse (August - March) show, that in order to obtain reliable equations, time series can be formed, including the monthly average values of chloroform content and water quality parameters for the entire observation period. Distinguishing two periods in an annual cycle provides acceptable estimates when modeling chloroform content. The period of low water and permanent watercourse can be described by regression equations characterized by a smaller average approximation error and a large value of the correlation coefficient. According to the results, a significant stochasticity in the annual period is introduced by April, which is the beginning of an active flood period. In general, the results obtained make it possible to consider the use of statistical modeling techniques to assess the value of the chloroform content by indicators of the quality of water in a water source - turbidity, chromaticity, and oxidizability.

The work was performed within the framework of the Ministry of Science and Higher Education of the Russian Federation state task in the field of scientific activity, publication number FEUR - 2020 - 0004 «Solving urgent problems and researching processes in petrochemical industries accompanied by flows of multiphase media».

References

1. M. A. Malkova, M. Y. Vozhdaeva, E. A. Kantor, IOP Conference Series: Materials Science and Engineering **775(1)**, 012092 (2020)

2. M. A. Malkova, M. Y. Vozhdaeva, E. A. Kantor, I. A. Melnitskii, IOP Conference Series: Earth and Environmental Science **315(5)** 052052 (2019)
3. R. Khan, I. Hashmi, A. Qureshi, S. Rasheed, Water and Environment Journal **35(1)**, 269-284 (2021)
4. K. Özdemir, Environment Protection Engineering **47(1)**, 87-97 (2021)
5. L. Godo-Pla, P. Emiliano, M. Poch, F. Valero, H. Monclús, Science of the Total Environment **763**, 144197 (2021)
6. R. A. Li, J. A. McDonald, A. Sathasivan, S. J. Khan, Water Research **190**, 116712 (2021)
7. A. Kennedy, L. Flint, A. Aligata, C. Hoffman, M. Arias-Paić, Water Research **188**, 116523 (2021)
8. G. Cool, I. Delpla, P. Gagnon, R. Sadiq, M. J. Rodriguez, Environmental Modelling and Software **120**, 104479 (2019)
9. M. Ramasamy, S. Nagan, Kumar P Senthil, Chemosphere **286**, 131571 (2021)
10. J. Tehrani et al., Science of the Total Environment **777**, 146223 (2021)
11. M. Yan, G-H. Fang, L-H. Dai, Q-F. Tan, X-F. Huang, Journal of Hydrology **600**, 126647 (2021)
12. P. M. Linnik, Hydrobiological Journal **57(1)**, 78-94 (2021)
13. N. A. Dilman, A. V. Mastryukova, S. E. Bednaruk, V. V. Chukanov, Environmental engineering **2**, 69-73 (2015)
14. A. Yu. Skryabin, G. V. Popovyan, I. A. Tron, Water supply and sanitary equipment **4**, 5-8 (2017)
15. V. A. Gushchin, A. I. Sukhinov, A. E. Chistyakov, A. V. Nikitina, A. A. Semenyakina, Computational Mathematics and Mathematical Physics **58(8)**, 1316-1333 (2018)
16. M. A. Malkova, A. I. Vasileva, N. D. Minchenkov, E. A. Kantor, IOP Conference Series Earth and Environmental Science **938(1)**, 012008 (2021)