The Prediction of the Trihalogenomethanes Content in Drinking Water: Infiltration Water Intake

M A Malkova¹, I I Beloliptsev² and E A Kantor¹

¹ Ufa State Petroleum Technological University, 1 Kosmonavtov str., Ufa 450064 Russia
² Financial University under the Government of the Russian Federation, 69 building 1, Mustaya Karima str., Ufa 450015 Russia

E-mail: kykyshka2009@mail.ru

Abstract. This study aims to establish a connection between the trihalogenomethanes content in drinking water of an infiltration intake and parameters characterizing water quality (turbidity, chromaticity, oxidizability), chlorine doses, and water consumption. Trihalomethanes include chloroform, bromodichloromethane, dibromochloromethane, and bromoform. It is known that trihalomethanes can harm public health. These compounds have a toxic polymorphism and can cause long-term effects, including carcinogenesis. It is shown that a satisfactory description of the trihalogenomethanes content can be obtained by modeling the annual cycle, taking into account the displacement of parameters (2–3 months), relative to the trihalogenomethanes content. Equations considering the displacement of water quality indicators have a high determination coefficient (0.95–0.97). The use of the same displacement for the time series of true values allows us to obtain equations with a sufficiently high determination coefficient, which indicates the possibility of their use for the predictive assessment of the trihalomethanes content in drinking water.

Keywords: Trihalogenomethanes · Turbidity · Chromaticity · Oxidizability · Water consumption · Prediction · Correlation Regression Analysis · Infiltration Water Intake

1. Introduction
One of the sanitary and epidemiological well-being measures is to provide the population with high-quality drinking water [1]. The water supply of urban agglomeration is usually organized through the use of surface or infiltration water. Infiltration water intake can provide higher drinking water quality. Water filtered through the soil is well clarified, since suspended particles are trapped in the soil pores. However, in this case, the stage of disinfection is required, especially if the water distribution networks are quite long [8].

Thus, water disinfection includes the complex of measures taken to prevent and eliminate infectious diseases transmitted by water [15, 16]. Chlorination is the most common and reliable method of water disinfecting.

A significant drawback of this method is the formation of chlorination products – organohalogen compounds of various compositions [2], most of which are trihalogenomethanes. This group of compounds includes chloroform, bromodichloromethane, dibromochloromethane, and bromoform. Trihalomethanes [THM] can harm public health. These compounds have a toxic polymorphism and can cause long-term effects, including carcinogenesis [3, 6, 7].
The sources of THM formation are anthropogenic and natural substances [3]. One of the most significant THM formation sources is the presence of aqueous humus in the river water [4, 9, 14] because, despite the filtration of groundwater through the soil, some parts of the humus penetrate through the filter layers.

Such factors also determine the quantitative THM content as temperature, the composition of organic compounds, the material of water pipes, the nature and dose of the chlorinating agent, pH, and others [5, 10, 11].

Analytical control of water quality at water treatment facilities includes turbidity, chromaticity, oxidizability, and does not provide for direct determination of THM precursors. In this regard, it seems reasonable to consider the relationship between the generalized indicators of water quality (turbidity, chromaticity, oxidizability), chlorine dose, and the amount of produced THM. Another parameter that can affect the concentration of THM in drinking water is water consumption. It largely determines the magnitude of the generalized water quality indicators due to the disruption of bottom sediments, washes from the coastal zone, and other reasons [12, 13].

2. Materials and Methods

The calculations were based on analytical control of the THM content in drinking water obtained by the Water Quality Analytical Control Center. The analytical control was performed in a water supply organization of a large urban agglomeration with infiltration-type water intake.

Multivariate correlation and regression analysis were used for modeling. This method allows us to identify the relationship between the THM content in drinking water with a chlorine dose (D\textsubscript{Cl}) and some generalized quality indicators of the water source: turbidity (T), chromaticity (C), oxidizability (O), which were supplemented by the flow rate of the incoming water during the analysis of the water source (Q).

The measurement of the THM content in drinking water is carried out once a month. In the year, there are 12 measurements. The total number of observations of the THM content was n = 216. In contrast to THM, turbidity, chromaticity, oxidizability, and water consumption at hydroelectric power stations are measured daily. Water consumption is the product of the average flow rate of the cross-sectional area of the watercourse. For these indicators, the average monthly values of water quality indicators were calculated by summing all values for the month and dividing the resulting value by the number of measurements in the month. Thus, data series were formed containing the same number of measurements as in the series of THM values (n = 216). The dose of chlorine was calculated based on the data of the water supply average monthly value per day (m\textsuperscript{3}/day) and the average monthly amount of chlorine spent (kg/day). As a result, the average values of the chlorine dose for each month were obtained.

According to the values of the monthly average dependences of the THM content, turbidity, chromaticity, oxidizability, water consumption, the average long-term values of these indicators are formed. This allows one to get 12 values of these indicators.

The described algorithm allows us to exclude from the study the influence of variable seasonal factors while maintaining the general tendency and dynamics of changes in the parameters of turbidity, chromaticity, oxidation, and flow rate of water entering the water source.

3. Results and Discussion

The main parameter that affects the THM content is the chlorine dose (D\textsubscript{Cl}). During the study (1996–2016), there was an increase in the trend of changes in the chlorine dose and THM content (figure 1).
Figure 1. Time series of the chlorine dose and THM content in drinking water of an infiltration intake. 

Source: Compiled by the authors.

\[
[\text{THM}] = b_{01} + b_{11}X; \tag{1}
\]

\[
[D_{\text{Cl}}] = b_{02} + b_{12}X \tag{2}
\]

where

[THM] – the THM content, \(\mu g/dm^3\);

[D_{\text{Cl}}] – chlorine dose, mg/dm^3;

\(b_{01}, b_{02}, b_{11}, b_{12}\) – coefficients of the regression equation;

\(X\) – the measurement number;

\(R^2\) – determination coefficient;

\(b_{01} = 6.845; \ b_{11} = 0.004; \ R^2(1) = 0.006; \ b_{02} = 0.432; \ b_{12} = 0.001; \ R^2(2) = 0.058\)

The results of calculating the correlation coefficient between the chlorine dose and the THM content show the presence of a “weak” bond strength on the Cheddock scale \((r = 0.10)\): 

\[
[\text{THM}] = b_{03} + b_{13} [D_{\text{Cl}}] \tag{3}
\]

where

\(r\) – correlation coefficient;

\(F\) – Fisher test;

\(b_{03}, b_{13}\) – the coefficients of the regression equation;

\(S\) – Student’s criterion;

\(A\) – relative error (%)

\(b_{03} = 4.9; \ b_{13} = 2.7; \ r = 0.1; \ R^2 = 0.01; \ F = 6.33; \ S = 22.2; \ A = 37.1\)

Even though the chlorine dose is a priority factor in the THM formation, modeling the THM content as a parameter, depending on the chlorine dose [5], does not allow satisfactory results. In this regard, an attempt was made to use multivariate correlation and regression analysis to simulate the THM content in drinking water obtained at the infiltration water intake. As independent variables, in addition to the chlorine dose, generalized indicators of water source water quality were used: turbidity, chromaticity, and oxidizability. These indicators are primarily related to the content of THM precursors in water, such as humus and fulvic acid. They determine the technology of water purification [9].

However, even in this case, it is impossible to obtain a satisfactory ratio between the THM content and the chlorine dose, turbidity, chromaticity, and oxidizability (table 1). It is interesting to note that chromaticity is an insignificant parameter of the regression equation (4).
Table 1. The multiple regression equation of the relationship between the THM content and the parameters: chlorine dose \([D_{Cl}]\), turbidity \([T]\), chromaticity \([C]\), oxidizability \([O]\).

| Coefficient \(b\) | designation | \(b_{04}\) | \(b_{14}\) | \(b_{24}\) | \(b_{34}\) | \(b_{44}\) |
|-------------------|-------------|-----------|-----------|-----------|-----------|-----------|
| value             | -0.68      | 10.46     | -0.11     | i/p*      | 0.85      |

| Equation characteristic | \(R^2\) | \(F\) | \(S\) | \(A\), % |
|-------------------------|--------|------|------|----------|
| Value                   | 0.08   | 6.33 | 5.7  | 67.0     |

*i/p*– insignificant parameter of the equation. Source: Compiled by the authors.

According to the equation, the chlorine dose and oxidizability increase the THM concentration, while the turbidity decreases. The value of the Fisher test indicates the statistical significance of the determination coefficient and the statistical reliability of the regression equation (i.e., the coefficients are jointly significant). However, the low value of the determination coefficient of the regression equation \((R^2 = 0.08)\) indicates an unsatisfactory description of the change in the THM content curve with the parameters (chlorine dose, turbidity, oxidizability). The average approximation error, using the regression equation (3), is 67.0%.

The river flow, on which the infiltration water intake is located, is regulated by the reservoir. Reservoirs have a significant impact on the river water regime and, therefore, on the river water’s qualitative composition. In turn, the water source’s quality depends significantly on the flow rate and quality of the river water. The introduction of the water flow rate through the water source can increase the degree of correlation between the THM content and parameters. In this regard, the equation (4), taking into account the chlorine dose, turbidity, chromaticity, and oxidizability, introduced an additional parameter – water consumption \((Q)\).

A comparative analysis of the obtained equations characteristics (4) and (5) shows that the introduction of a water consumption parameter into the regression equation has little effect on the features of the bond strength between the THM content and the parameters (table 3). The value of the determination coefficient increased slightly, and the relative error was practically preserved (table 2).

Table 2. The multiple regression equation of the relationship between the THM content and the parameters: chlorine dose \([D_{Cl}]\), turbidity \([T]\), chromaticity \([C]\), oxidizability \([O]\), water consumption \([Q]\).

| Coefficient \(b\) | designation | \(b_{05}\) | \(b_{15}\) | \(b_{25}\) | \(b_{35}\) | \(b_{45}\) | \(b_{55}\) |
|-------------------|-------------|-----------|-----------|-----------|-----------|-----------|-----------|
| value             | -27.43      | 41.99     | -0.96     | -0.56     | 3.98      | 0.03      |

| Equation characteristic | \(R^2\) | \(F\) | \(S\) | \(A\), % |
|-------------------------|--------|------|------|----------|
| Value                   | 0.20   | 8.7  | 7.7  | 63.5     |

Source: Compiled by the authors.

Thus, the use of the regression equation (5) is not possible.

It can be assumed that the reason for the weak relationship between the THM content and the chlorine dose, water consumption, and parameters of its condition is the influence of climatic conditions, such as, for example, different times of the spring beginning and autumn floods. Some averaging of such an
effect can be achieved by convolving the observed values of the analyzed number series — instead of 216 values of the number series, we get 12. In this case, each of the 12 values will correspond to a value showing the system’s state in a particular month. The obtained data is a description of the annual cycle, which is the average for the entire observation period (figure 2).

Figure 2. Annual time series of changes in the THM content, chlorine dose, chromaticity, oxidizability, turbidity, and water consumption of the infiltration water intake. Source: Compiled by the authors.

The regression equation indicators, formed according to the average annual cycle (table 3), indicate the possibility of using equation 6 to assess the quality of drinking water according to the THM content. Thus, the correlation coefficient is 0.89, and the approximation error is 31.2%.

Table 3. The multiple regression equation of the relationship between the THM content and the parameters: chlorine dose \([D_{Cl}]\), turbidity \([T]\), chromaticity \([C]\), oxidizability \([O]\), and water consumption \([Q]\) according to the parameters of the annual cycle.

| The equation \([THM] = b_{06} + b_{16}[D_{Cl}] + b_{26}[M] + b_{36}[O] + b_{46}[Q]\) | (6) |
|---|---|
| Coefficient \(b\) | designation | \(b_{06}\) | \(b_{16}\) | \(b_{26}\) | \(b_{36}\) | \(b_{46}\) | \(b_{56}\) |
| value | \(-27.43\) | \(121.00\) | \(-2.49\) | \(-1.78\) | \(16.22\) | \(0.05\) |
| Equation characteristic | \(R^2\) | \(F\) | \(S\) | \(\Lambda\), % |
| Value | 0.79 | 6.2 | 2.2 | 31.2 |

Source: Compiled by the authors.

It is noteworthy that when constructing changes in the THM content and other parameters in the annual cycle, a displacement in their maximum values, relative to each other, is observed. Thus, for THM, the maximum concentration occurs in July and coincides with the chlorine dose for chromaticity, oxidizability – in May, for turbidity – in April (Fig. 2). Thus, the THM content is delayed by 2–3 months compared with the maximum concentrations of chromaticity, oxidizability, water consumption, and turbidity.

Multivariate regression analysis performed to find a relationship between the THM content and parameters showed a significant increase in the correlation coefficient \((r = 0.98)\) and a decrease in relative error (table 4).
The obtained regression equation (7) for the THM content forecasting in the annual cycle (long-term forecasting), taking into account the displacement in turbidity, chromaticity, oxidizability, and river water consumption, is reliable. It has a high determination coefficient; the average relative forecast error was 5.8% (Table 5). It seems interesting to analyze the possibility of modeling the annual cycle, using displacement, the time series of water parameters (turbidity, chromaticity, oxidizability, and water flow), and the true time series for the entire observation period. For this purpose, a regression equation is defined. It describes the period 2011–2014 (Table 6).
Table 6. The equation of multiple regression of the relationship between the content of THM and the parameters of the chlorine dose [Dₐₙ], turbidity [T], chromaticity [C], oxidizability [O], and water consumption [Q], taking into account the displacement of the parameters maximum values relative to the THM content in the period 2011–2014. (displacement by turbidity – 3 months, color, oxidizability, and consumption – 2 months, chlorine dose – without displacement).

| Coefficient b designation | b₀₈ | b₁₈ | b₂₇ | b₃₈ | b₄₈ | b₅₈ |
|---------------------------|-----|-----|-----|-----|-----|-----|
| value                     | -11.76 | 24.02 | 0.08 | 0.05 | -0.32 | 0.01 |

Equation characteristic

\[ \text{R}^2 = 0.65 \]
\[ F = 15.81 \]
\[ S = 6.3 \]
\[ A, \% = 42.13 \]

Source: Compiled by the authors.

We can assume that, in this case, the necessary reliability of the equation is achieved, and it becomes possible to simulate the THM content not only in the annual cycle but also for the current time (figure 3).

Figure 3. Actual and forecasted THM concentrations in the period 2012–2015. Source: Compiled by the authors.

Because the THM analysis accuracy is 26%–42%, the obtained equation can be considered acceptable for a preliminary assessment of the THM content in drinking water.

4. Conclusion

The analysis of the THM predicting possibility, depending on the water quality, generalized indicators of the water source (turbidity, chromaticity, oxidizability), the chlorine dose, and consumption, shows that the THM content predicting in drinking water is marked with a low determination coefficient.

Using the average monthly values of the specified parameters for predicting the increase parameters connection degree. This allows using the obtained regression equations for the long-term prediction of the THM concentration. Additionally, a comparison of the monthly average values of turbidity, chromaticity, oxidizability, chlorine dose, water flow, and THM concentrations revealed that the maximum THM concentrations were shifted to others in 1–3 months. Equations considering the displacement of water quality indicators have a high coefficient of determination (0.95–0.97). The application of the same displacement for a time series of true concentrations allows obtaining equations with the determination coefficient \( \text{R}^2 = 0.65 \).

5. Acknowledgments

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] Ademollo N, Patrolecco L, Polesello S, Valsecchi S, Wollgast J, Mariani G, ... Hanke G 2012 The analytical problem of measuring total concentrations of organic pollutants in whole water Trends Anal. Chem. 36 pp 71-81

[2] Bond T, Goslan E H, Parsons S A, and Jefferson B 2012 A critical review of trihalomethane and haloacetic acid formation from natural organic matter surrogates Environmental Technology Reviews 1(1) pp 93-113

[3] Bulla R, Reckhow D, Li X, Humpeaged A, Joll C, and Hrudec S 2001 Potential carcinogenic hazards of non-regulated disinfection by-products: Haloquinones, halo-cyclopentene and cyclohexene derivatives, N-halamines, halonitriles, and heterocyclic amines Toxicology 286 pp 1-19

[4] El-Shafy M A, and Grunwald A 2000 THM formation in water supply in South Bohemia, Czech Republic Water Research 34(13) pp 3453-3459

[5] Goncharuk V V 2014 Drinking water: Physics, chemistry, and biology (Cham, Switzerland: Springer International Publishing)

[6] Hrudey S E, Backer L C, Humpeaged A R, Krasner S W, Michaud D S, Moore L E, Singer P C, … Stanford B D 2015 Evidence for Association of Human Bladder Cancer with drinking-water chlorination disinfection by-products. Journal of Toxicology and Environmental Health Part B CriticalReviews 18(5) pp 213-241

[7] Malkova M A, Vozhdaeva M Y, Kantor E A, and Melnitski I A 2019 Comparison of drinking water quality in terms of relative concentrations and carcinogenic risk of trihalomethanes IOP Conf. Ser.: Earth Environ. 315 052052

[8] Nieuwenhuijsen M J, Frellier J, Smith R, Iszatt N, Bennett J, and Best N 2009 The epidemiology and possible mechanisms of disinfection by-products in drinking water Philos Transact A Math PhysEng Sci. 367 (1904) pp 4043-4076

[9] Pereira J, Moreira C, Albers C, Jacobsen O, and Flores E 2011 Determination of total organic halogen (TOX) in humic acid after microwave-induced combustion Chemosphere 83 pp 281-286

[10] Platikanov S, Puig X, Martin J, and Tauler R 2007 Chemometric modeling and prediction of trihalomethane formation in Barcelona’s waterworks plant Water Research 41(15) pp 3394-3406

[11] Ricci M, Kourchev I, and Emons H 2012 Chemical water monitoring under the water framework directive and certified reference Materials Trends Anal. Chem. 36 pp 47-57

[12] Rodriguez M J, Vinette Ya, Serodes J B, and Bouchard Ch. 2003 Trihalomethanes in drinking water of the greater Quebec region (Canada): occurrence, variations and modeling Environmental Monitoring and Assessment 89(1) pp 69-93

[13] Shan J, Hu J, Kaplan-Bekaroglu S, Song H, and Karanfil T 2012 The effects of pH, bromide, and nitrate on halonitromethane and trihalomethane formation from amino acids and amino sugars Chemosphere 86 pp 323-328

[14] Sun Y-X, Wu O-Y, Hu H-Y, and Tian J 2009 Effect of bromide on the formation of disinfection by-products during wastewater chlorination Water Res. 43 pp 2391–8

[15] Virtius E, Kapinus E, Revelsky I, and Borzenko A 2003 Determination of total organochlorine content in water, based on microfluid extraction and microcoulometric analysis of the extract Zav. Lab 69 pp 3-6

[16] Vozhdaeva M Yu, Tsypysheva L G, Kantor L I, and Kantor E A 2015 Influence of chlorination on the composition of partly volatile organic contaminants of water Russian Journal of Applied Chemistry 77(6) pp 938-941