Assessment of trends in chloroform content in drinking water of infiltration water intake

Maria Malkova¹ *, Olga Kantor¹, Alina Yalaletdinova¹, and Evgeny Kantor¹

¹Ufa State Petroleum Technological University, 1, Kosmonavtov Street, Ufa, 450062, Russia

Abstract. Chloroform is a common result product of water disinfection. In drinking water, chloroform is present in higher concentrations than the other trihalogenmethanes (THM) (bromodichloromethane, dibromochloromethane and bromoform). For this reason, it is considered as an indicator of the content of a group of compounds related to THMs. It was found that the accumulation rate of chloroform based on the use of Markov chains in the medium and short term in drinking water increases over time.

1 Introduction

Chloroform is a representative of trihalogenmethanes (THM) group. Chloroform is the main product of the result of water disinfection with chlorine-containing compounds [1-3]. Complexity of the ways of chloroform entry into the body predetermines the appearance of increased risks for human health [4-6]. Thus, in domestic conditions, chloroform affects humans not only enterally, but also through the lungs due to its volatility [4-5]. In drinking water chloroform occurs in higher concentrations than other THMs (bromodichloromethane, dibromochloromethane and bromoform). For this reason, it is considered as an indicator of the content of chlorination products in water [1].

Water quality is influenced by a large number of factors. This allows us to consider the dynamics of water quality changes from the perspective of the theory of Markov random processes [7-8].

Markov processes are a convenient mathematical tool for studying objects of different nature in order to predict the dynamics of their changes in the medium and short-term periods [9-12]. A particular case of Markov processes are Markov chains, for which it is assumed that in each period of time the system can be in one of a finite or countable set of states, and the transition from one state to another (including itself) occurs at discrete moments of time. If the entire set of possible water quality characteristics is reduced to several states, taking into account the current practice of collecting information about the water source, modeling and forecasting of the studied water quality characteristics can be carried out on the basis of Markov chains [7-8].

* Corresponding author: malkova_maria92@vk.com

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).
2 Materials and methods

The data of analytical determination of chloroform (CF) content for a twenty-year period in drinking water obtained by the analytical control water quality centre of the large urban agglomeration water supplying organization at the infiltration-type water intake served as a baseline for calculations [13].

Approaches based on the application of mathematical and statistical methods do not always provide answers to all questions of interest. In such cases, taking into account the possibility of attributing water quality to the so-called non-sequential processes, modeling from the perspective of discrete Markov processes [7, 9-10] can become an effective tool for analyzing and predicting the quality characteristics of a water source. The application of Markov random processes implies setting a set of states, determining the initial state and setting the intensity matrix of transitions from one state to another [7].

Let us assume that at any time $t$, water quality can be differentiated into three non-intersecting states: $S_1$ - "satisfactory", $S_2$ - "acceptable" and $S_3$ - "unsatisfactory".

Each transition from state $S_i$ to state $S_j$ is characterized by a transition probability $p_{ij}$. Thus, the transition matrix has the form:

$$
\mathbf{P} = \begin{bmatrix}
    p_{11} & p_{12} & p_{13} \\
    p_{21} & p_{22} & p_{23} \\
    p_{31} & p_{32} & p_{33}
\end{bmatrix}
$$

(1)

For the elements of its rows, the relation is satisfied:

$$
p_{i1} + p_{i2} + p_{i3} = 1, \quad i = 1, 3
$$

(2)

Probabilities of staying characteristics of system elements in states $\{S_1, S_2, S_3\}$ at time $(t+1)$ for a homogeneous Markov chain are obtained from the relations:

$$
P(t+1) = P(t) \cdot \mathbf{P}
$$

(3)

Where $P(t) = \{P_i(t) | i = 1, 3\}$, $P(t) = \{P_{i1}(t+1) | i = 1, 3\}$, $P_i(t)$ and $P_{i1}(t+1)$ are the probabilities that water quality will be assigned to state $S_i$ at time $t$ and $(t+1)$, respectively.

At the initial moment of time, the vector $P(0)$ is formed according to the actual water quality (i.e., one of its components should be equal to 1 and the rest should be zeros). Then the vector $P(m)$ of probability distribution over the states at step $m$ will be determined from the relation:

$$
P(m) = P(0) \cdot P^m
$$

(4)

In practical calculations, the probabilities of transitions from one state to another and the probabilities $P_i(t)$ are initially unknown, so instead, relative frequencies calculated for the corresponding quantities are used.

When performing operations related to correlation and regression analysis, the following notations are accepted: $k$ - angular coefficient; $b$ - free coefficient; $R^2$ - coefficient of determination; $F$ - Fisher distribution statistic; $A$ - average relative approximation error.
3 Results

The content of CFs is of considerable interest from the point of view of drinking water quality management. CF can be detected in source water, but their content is usually less than 5% of the maximum permissible concentration (MPC) [13]. The result of the presence of CFs in drinking water is the reactions that occur during the action of hypochlorous acid or its salts on organic compounds contained in the source water [1-3]. It is believed that one of the main factors affecting the quantitative content of CFs is the dose of chlorine-containing reagent used for water disinfection [14].

Earlier studies were carried out to assess the relationship between the content of THM and its components with water quality indicators of the water source (turbidity, chromaticity, permanganate oxidizability), as well as with such parameters as chlorine dose and water flow rate of the water source [15-16]. The obtained data allow estimating the content of THM and CF in drinking water with a relative prediction error from 42% to 67% at different approaches in the search for the considered dependence [15-16]. However, a sufficiently good result can be obtained only with models that use the steps of averaging and shifting of the original (true) time series [15-16]. For example, when modeling monthly averages, the degree of relationship between the content of THM and its components in drinking water and turbidity, chromaticity and permanganate oxidizability of the water source, chlorine dose, and water flow rate is characterized by a multiple correlation coefficient of 0.80 for surface water intake and 0.97 for infiltration water intake, which allows using the obtained regression equations for long-term prediction of THM concentration [15-16]. Comparison of average monthly values of turbidity, chromaticity, permanganate acidification, chlorine dose and concentrations of THM components revealed that the maximum of THM and CF concentrations is shifted relative to others by 1 - 3 months. Correlation-regression equations, which take into account the displacement of water quality indicators of the water source in relation to the content of THM (CF), have a high value of the coefficient of determination (0.86 - 0.98). Application of the same bias for time series of true concentrations for both types of water intakes allows us to obtain equations with a coefficient of determination of 0.62 and 0.65 for surface and infiltration water intakes, respectively [15, 16].

It is important to note that the above studies used an approach related to the search for the dependence of THM and/or CF content on water quality parameters of the water source and other factors that somehow affect this quality [15, 16].

Predicting the content of THM (or CF) from the data of its concentration in a certain time period is possible using the method of time series analysis. However, in this case, the prediction error can reach the value of more than 150% [17].

Predicting the content of CFs in drinking water as an indicator of the content of chlorination products in it is complicated by the fact that the time series of THM and, in particular, CF concentration is described by a random component by more than 70% [18]. The share of the trend-cyclic component, as a rule, is less than 1%.

The characteristics of the obtained linear regression equation of the content of CFs in drinking water for the twenty-year period of observations indicate that it is not possible to obtain a satisfactory prediction of the content of CFs in drinking water with the help of linear regression models on the true concentrations (Table 1).

Table 1. Results of predicting chloroform content (µg/dm$^3$) using linear regression model over a twenty-year period (x is the ordinal number of the measurement).

<table>
<thead>
<tr>
<th>Equation</th>
<th>$A$, %</th>
<th>$R^2$</th>
<th>$F$</th>
<th>Forecast month</th>
</tr>
</thead>
<tbody>
<tr>
<td>(5) $CF = 0.0039 \cdot x + 4.25$</td>
<td>52.2</td>
<td>0.08</td>
<td>2.19</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.95</td>
</tr>
</tbody>
</table>
This is due to the fact that the initial data have a high degree of stochasticity, the share of the random component of the time series reaches 70%, and the model does not have the necessary level of validity and reliability. Thus, it can be considered that the potential of using such an approach for processing in-situ measurement data in this case is exhausted by revealing the general trend of changes in the CF content.

On the other hand, it is of considerable interest to determine the possibility of fluctuations in the values of THM and CF content in the short term. In this respect, Markov chains are of some interest [7-8].

Some data related to an attempt to analyze water quality conditions with respect to CFs with a 1-month forecast have been presented previously [20]. Evaluation of the probability of chloroform concentration hitting the states $S_1$, $S_2$, $S_3$ for the period equal to 1 month, when the variation of retrospective data depth from 20 years to 1 year showed that when the variation of retrospective data depth from 8 to 4 years, there is a sharp deterioration of the situation: the probability of hitting the values of CF concentration in the first region (state $S_1$) decreases from 0.77 to 0.57, and in the third region (state $S_3$) - increases from 0.05 to 0.14 [20].

As previously [20] for the realization of the steps determining the use of Markov chains, we have identified three ranges of chloroform concentrations in drinking water. The first range includes the area of chloroform concentrations from 0 to 6 $\mu$g/dm$^3$ and represents the zone of the lowest potential risks with respect to causing harm to public health (state $S_1$). The second range is from 6 to 10 $\mu$g/dm$^3$ (State $S_2$). This is the range of the most commonly encountered concentrations of CFCs in drinking water. The third range - greater than 10 $\mu$g/dm$^3$ - is characterized by an increased potential risk (State $S_3$). The transition probabilities $p_{ij}$ are determined from the available data.

Using finite Markov chains in accordance with the sequence of calculations described above, we predicted the probability of hitting the concentration of CF in the states $S_1$, $S_2$, $S_3$ for a period of up to 4 months at variation of the depth of retrospective data from 20 years to 1 year (Table 2).

Table 2. Estimated probability of chloroform concentrations in drinking water falling into ranges 1-3 for a period of 1 - 4 months.

<table>
<thead>
<tr>
<th>Range</th>
<th>Depth of retrospective data, years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20</td>
</tr>
<tr>
<td>The first month</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.84</td>
</tr>
<tr>
<td>2</td>
<td>0.14</td>
</tr>
<tr>
<td>3</td>
<td>0.02</td>
</tr>
<tr>
<td>Second month</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.80</td>
</tr>
<tr>
<td>2</td>
<td>0.18</td>
</tr>
<tr>
<td>3</td>
<td>0.02</td>
</tr>
<tr>
<td>Third month</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.77</td>
</tr>
<tr>
<td>2</td>
<td>0.20</td>
</tr>
<tr>
<td>3</td>
<td>0.03</td>
</tr>
<tr>
<td>Fourth month</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.76</td>
</tr>
<tr>
<td>2</td>
<td>0.20</td>
</tr>
<tr>
<td>3</td>
<td>0.04</td>
</tr>
</tbody>
</table>
Analysis of the obtained data shows that at the depth of retrospective data in the interval from 20 to 10 years the probabilities of attributing water quality to the states $S_1$, $S_2$, $S_3$ are practically unchanged. The share of results of analytical control of CF content increases significantly in the areas $S_2$ and $S_3$.

In general, this indicates a fairly significant deterioration in water quality. The increase in predicted CF concentrations is particularly noticeable for the retrospective period starting at four years for the first month. For months 2, 3, and 4, these values increase more smoothly (Figure 1).

![Fig. 1. Probability of change in the concentration of CF in drinking water when predicting its content for 1, 2, 3 and 4 months at different depth of retrospective data.](image)

**4 Conclusions**

Thus, Markov chains make it possible to assess the quality of water subjected to chlorination in retrospect. This mathematical tool is quite convenient for tracking changes in water quality, in particular, of CF, in operational and short-term periods, and, on the other hand, it can also be convenient for detecting disturbances not captured by analytical control related to events causing fluctuations of THM and CF concentrations in drinking water.

Thus, according to the results of the totality of conducted studies, it can be considered that linear regression models, in general, reflect the deterioration of drinking water quality by such indicator as the content of CFs. The application of Markov chains allows us to obtain data on changes in drinking water quality in the short and medium term [7-12], to reveal that for the forecast period (1 - 4 months) there is a marked increase in the rate of accumulation of CFs, which ultimately increases the risk of deterioration of drinking water quality.

In general, the combination of methods for analyzing long-term trends obtained using linear regression models and probabilistic estimates of the content of CFs, calculated on the basis of Markov chains, allows to quantitatively assess the dynamics of ongoing changes in drinking water quality from the position of the presence of chlorination products in it.

**Acknowledgments**

The work was carried out within the framework of the State Assignment of the Ministry of Science and Higher Education of the Russian Federation, publication number FEUR - 2023 - 0006, project "Development and creation of low-tonnage products and reagents (corrosion and salt deposition inhibitors, antioxidants, biocides, additives, etc.) for oil and gas chemistry processes and water purification from contaminants, replacing imported substances and materials. Theoretical and experimental approaches".
References


5. B.I. Marchenko, P.V. Zhuravlev, N.K. A.I. Yuhno, Assessment of carcinogenic risk from exposure to chlororganic compounds of water of systems of centralized water supply. Gigiena i Sanitariya, 100(2) (2021)


10. O.V. Admaev, T.V. Gavrilenko, Expert system of atmospheric pollution assessment as a tool of environmental control and technical means of observation and monitoring of the environment. Mining information-analytical bulletin (scientific and technical journal), 18(12) (2009)

11. A.A. Malashonok, M.O. Singegovsky, Simulation and Forecasting of Soybean Yield in Amur Region. Dostizheniya nauki i tehniki APK, 31(8) (2017)

