Probabilistic assessment of water quality in terms of oxidability

A V Yalaletdinova¹, E A Kantor², I I Beloliptsev³ and Y O Galimova⁴

¹ Department of “Physics”, Ufa State Petroleum Technological University, 1 Kosmonavtov St, Ufa, 450062, Russia
² Department of “Physics”, Ufa State Petroleum Technological University, 1 Kosmonavtov St, Ufa, 450062, Russia
³ Department of "Mathematics and Informatics", Financial University under the Government of the Russian Federation, Ufa Branch, Mustaja Karima st. 69/1, Ufa, 450015, Russia
⁴ National Library of Bashkortostan, Lenina st. 4, Ufa, 450077, Russia

E-mail: hawk22.89@mail.ru

Abstract. There has been analyzed the nature of the distribution of the oxidability of water. It is revealed that the nature of the distribution of indicator values during the year depends largely on the seasonal factor, and therefore the analysis of the distribution of oxidability is proposed to be studied separately for each month. A variation series is constructed and empirical distribution functions of oxidability distribution is derived. It is established that the law of water oxidability distribution differs from the normal and lognormal distributions, but it is described with sufficient accuracy by the gamma distribution or by a cubic polynomial function (being the simplest). The hypothesis about the distribution law is confirmed by the Kolmogorov–Smirnov test. The water oxidability distribution function allows to determine the probability of exceeding the specified values of the indicator and quantitatively assess the risks of exceeding them, which can become the basis for developing solutions for managing water quality and increasing the efficiency of the water treatment process.

1. Introduction

Natural water is a multiphase environment which contains various impurities. Their quantitative and qualitative composition depends on the influence of many factors, but the determining factors are the location of the water source and the type of water intake facility [1-3]. Surface water intakes are more affected by natural, anthropogenic, and technology-related factors [1], compared to infiltration ones. In metropolitan areas with developed oil refining, petrochemical, engineering and construction industries, the impact of anthropogenic factors on water resources contributes to the fact that the indicators that are the main technological parameters during the organization of water treatment (which include oxidability) are determined not only by significant annual climatic fluctuations [2, 3], but also by man-caused impact from enterprises that are sources of pollutants of various nature [4, 5].

Oxidability is a quality composite index and characterizes the presence of both organic and nonorganic components in water. The components are suspended matter, colloidal matter, and truly dissolved solids which undergo various stages of transformation and redistribution during migration.
For example, it was found that the composition of oxidizability includes organic matter of industry-related origin (Benz(a)pyrene, which is a global pollutant) [6, 7], as well as high-density metals (copper, zinc, lead) absorbed by organic matter [8]. To improve the efficiency of the water treatment process, water quality control is necessary, particularly during periods characterized by deterioration of certain indicators [9], for example, oxidizability, which is the main parameter reflecting the presence of organic matter in water [9, 10].

Therefore, it is important to have a symbolic model that describes the seasonal distribution of water oxidizability (the probability of occurrence of situations in which the indicator might exceed or not exceed the predetermined values), since it can be used to form descriptions of the reagent dose, flow rate control, and solutions to other problems related to improving water treatment technology.

2. Research objectives and methodology

The objects of research were the data of daily measurements of water oxidizability for the period from 1997-2014 in the gauge of surface water intake.

To build a symbolic model, statistical analysis was performed, which involved determining the relative frequency polygon of oxidizability susceptibility, finding an empirical distribution function, and testing hypotheses about the theoretical distribution law [3].

The relative frequency polygon is constructed from the original series of oxidizability data. The range of values of the indicator is divided into intervals of length k, the boundaries of each interval are marked \((x_{\text{min}}, x_{\text{max}})\), the frequencies \((n_i)\) and relative frequencies \((w_i = n_i/n)\) of values falling into the \(i\) interval are determined.

As an empirical distribution function \(F_n(x)\), we consider a number of values of the accumulated relative frequency obtained from experimental data. For all intervals, the values of the accumulated relative frequency \(F_n(x)\) are determined cumulatively and lie in the range from 0 to 1. This way, the values of \(F_n(x)\) are interpreted as the empirical probability that the values of the random variable \(X\) will not exceed the values of the argument \(x\): \(F(x) = P(X < x)\). The theoretical distribution function \(F^2(x)\) is found by approximating a series of accumulated relative frequencies \(F_n(x)\) by some function [3].

Hypotheses about the correspondence of the empirical distribution of water oxidizability to the normal and lognormal laws of probability distribution [11] were tested using the Pearson’s \(\chi^2\) criterion [12].

The gamma distribution [11] of water oxidizability is given by the distribution density function of the form

\[
f(x) = \begin{cases} 
  b^a \frac{x^{a-1} e^{-bx}}{\Gamma(a)}, & 0 \leq x < \infty \\
  0, & x < 0
\end{cases}
\]

where \(a\) and \(b\) are the parameters of the gamma distribution; \(\Gamma(a)\) - the Euler gamma function.

The parameters of the gamma distribution are calculated as:

\[
a = \frac{\bar{x}^2}{s^2}, \quad b = \frac{\bar{x}}{s^2},
\]

where \(\bar{x}\) is the sample mean; \(s^2\) is the sample variance [3].

The hypothesis \((H_0)\) about the correspondence of the empirical and theoretical distributions of oxidizability is checked by the Kolmogorov test criterion [12], for this purpose, a measure of deviation between the theoretical and empirical distributions are calculated:

\[
D = \max \left| F_n(x) - F(x) \right|
\]

Next is the value \(\lambda = Dn^{1/2}\), where \(n\) is the sample size. For the selected significance level \(\alpha = 0.05\), the table value is \(\lambda_\alpha\) (\(\lambda_{0.05} = 1.36\)). The hypothesis \(H_0\) will not contradict the experimental data if \(\lambda \leq \lambda_\alpha\).
The domain of definition of the theoretical distribution function \( F'(x) \), where all properties of cumulative distribution function [3, 12] will be performed, is assumed from the condition:

\[
x \in [x_1; x_2] \subseteq [x_{\text{min}}; x_{\text{max}}],
\]

where \( x_1, x_2 \) - argument values at which the distribution function of the random variable \( F'(x) \) is continuous and monotonically increasing on the segment \([0;1]\); \( x_{\text{min}}, x_{\text{max}} \) - the minimum and maximum values of the sample.

3. Results and discussion

The sample size \((n)\) for water oxidation in the river for the period from 01.01.1997 to 31.12.2014 (daily observations) amounts to 6574 values.

In April - May, an increase in water discharge in the river leads to a sharp increase in oxidability. The values of the indicator change not only under the influence of the flow rate that causes the erosion of bottom sediments, but also as a result of humic substances washout from the terrain [9, 10]. Therefore, the oxidability of water is considered as a random variable.

To analyze the distribution of oxidability over the entire record period, relative frequency polygon is built according to the initial sample \( n \) (figure 1).

![Figure 1. Range of relative frequencies of oxidability from 1997 to 2014.](image)

About 90% of the records fall in the first interval (up to 3.5 mgO/dm\(^3\)), the remaining observations form a long right tail of the distribution, which estimates the probability of occurrence of extreme values of oxidability. When studying the distribution of the indicator for the entire record period, seasonal features are not taken into account: large figures of oxidability are recorded during floods and fluctuating ones within a certain interval can be seen in other, more stable periods. Thus, the appearance of high oxidability (\( O \geq 5 \) mgO/dm\(^3\)) in January or September is an almost impossible event and its probability tends to zero (figure 2).

![Figure 2. Average monthly oxidability values 2011-2014.](image)

That is why the law of oxidability distribution is studied for each month separately. For a detailed consideration of the nature of the distribution of water oxidability, we selected January as a month characterized by small values of the indicator and May, when the oxidability exceeds its MPC (maximum permissible concentration) (figure 2).

Indicator values from the original sample \( n \) are grouped for each month. The sample sizes \((n_i)\) of each month are determined, as well as the minimum \((x_{\text{min}})\) and maximum \((x_{\text{max}})\) values of oxidability for each sample, and a set of variate values is built, under which the width of the interval is taken \( \Delta S \): for January - \( n_i = 553, x_{\text{min}} = 0.6 \) mgO/dm\(^3\), \( x_{\text{max}} = 4.2 \) mgO/dm\(^3\), \( \Delta S = 0.4 \); for May - \( n_i = 533, x_{\text{min}} = 2.7 \) mgO/dm\(^3\), \( x_{\text{max}} = 10.4 \) mgO/dm\(^3\), \( \Delta S = 1.5 \).
Interval boundaries are set for each sample, and relative frequencies \(w_i\) are calculated (figure 3) as well as the values of the empirical distribution function \(F_n(x)\) (table 1).

\[
\begin{align*}
\text{Figure 3.} & \quad \text{The relative frequency range} \; o_i \text{ and gamma distribution function curve} \; f(x) \\
& \text{of oxidability from 1997 to 2014:} \; a) \; \text{January,} \; b) \; \text{May.}
\end{align*}
\]

**Table 1.** Characteristics of empirical, gamma, and theoretical distributions of water oxidability: \(\hat{F}(x)\) - cumulative frequency, \(F(s)\) - values of the gamma distribution function, \(F(s)^1\) - values of the distribution function.

| \(\hat{F}(x)\) | \(F(s)\) |
|----------------|----------|
| 0.005          | 0.005    |
| 0.001          | 0.003    |
| 0.997          | 0.998    |

Hypotheses about the correspondence of empirical distributions to logarithmic normal distribution, under Pearson’s criterion \(\chi^2\) [11], are denied.

Relative frequency polygons (figure 3) provide the possibility that the water oxidability has gamma distribution [11]. Parameters \(a\) and \(b\) of distributions (1) and values of the density function of the gamma distribution (table 1) are determined by substituting in (2) the values of the sample mean \(\bar{x}\) and sample variance \(s^2\): for January - \(\bar{x} = 1.973, \; s^2 = 0.615, \; a = 4.814, \; b = 2.444\); for May - \(\bar{x} = 6.123, \; s^2 = 1.897, \; a = 19.766, \; b = 3.228\).

The empirical data do not completely coincide with the selected theoretical distributions (figure 3). The hypothesis about the gamma distribution of water oxidability is checked by Kolmogorov test [12] (table 2).
For January, the values $D = 0.051$, $\lambda = 1.19$. Tabulated point for the significance level $\alpha = 0.05$ is equal to $\lambda_{0.05} = 1.36$. As far as $\lambda \leq \lambda_{0}$, the hypothesis that the distribution of water oxidation in January has a gamma distribution with parameters $a = 4.814$; $b = 2.444$, is adopted. For May the varieties $D = 0.021$, $\lambda = 0.50$. As far as $\lambda \leq \lambda_{0}$, the hypothesis that the distribution of water oxidation in May has a gamma distribution with parameters $a = 19.766$; $b = 3.228$, is adopted.

**Table 2. Results of testing the hypothesis about the distribution law by Kolmogorov test.**

|     | JanuarY |          |          |          |          | OctoBer |
|-----|---------|----------|----------|----------|----------|---------|
| $s_{2}$ |         | 0.250    | 0.423    | 0.539    | 0.675    | 0.857   | 0.955   | 0.982   | 0.998   | 1.000   |
| $F_{d}(s_{2})$ | 0.034 | 0.051    | 0.036    | 0.051    | 0.024    | 0.051   | 0.035   | 0.027   | 0.015   |          |

**Results of testing the hypothesis about the distribution law (1)**

|     | JanuarY |          |          |          |          | OctoBer |
|-----|---------|----------|----------|----------|----------|---------|
| $s_{2}$ |         | 0.228    | 0.402    | 0.565    | 0.709    | 0.832   | 0.926   | 0.988   | 1.011   | 0.990   |
| $F_{d}(s_{2})$ | 0.005 | 0.125    | 0.515    | 0.837    | 0.975    | 1.000   |          |          |          |

**Results of testing the hypothesis about the distribution law (2)**

|     | JanuarY |          |          |          |          | OctoBer |
|-----|---------|----------|----------|----------|----------|---------|
| $s_{2}$ |         | -0.171   | 0.487    | 0.814    | 1.026    | 0.998   |          |          |          |
| $F_{d}(s_{2})$ | 0.015 | 0.046    | 0.029    | 0.024    | 0.051    | 0.002   |          |          |          |

There was also done the polynomial approximation of amounts of empiric distribution function $F_{d}(x)$ (figure 4):

January: $F'(x) = -0.0139x^{3} + 0.0297x^{2} + 0.4341x - 0.3113$, $R^{2} = 0.9957$ (5)

May: $F'(x) = -0.0062x^{3} + 0.114x^{2} - 0.4701x + 0.5426$, $R^{2} = 0.9935$ (6)

**Figure 4.** Empirical $F_{d}(x)$ and theoretical $F'(x)$ functions of oxidability distribution: a) January, b) May.

The region of acceptable values of argument $x$ of functions (5, 6), at which the functions have all properties of the distribution function (positive and increase monotonically), is formed from condition (4):

January: $x \in [0.695; 3.733] \subset [0.6; 4.2] = [0.695; 3.733]$ (7)
Over the entire long period of records (from 1997 to 2014), the oxidability has never exceeded $x_{\text{max}}$, though the probability that the indicator will exceed the specified value exists, but it is highly unlikely. That is why, in accordance with the principle of practical confidence, we consider the probability that oxidability will not exceed $x_{\text{max}}$ equal to 1. Similarly, the occurrence of values less than $x_{\text{min}}$ - we consider almost an impossible event with a probability equal to zero:

\[
\begin{align*}
    &\text{January:} & F(x) &= \begin{cases} 
        0; & x < 0.695 \\ 
        -0.0139 \cdot x^3 + 0.0297 \cdot x^2 + 0.4341 \cdot x - 0.3113; & 0.695 \leq x \leq 3.733 \\ 
        1; & x > 3.733 
    \end{cases} \\
    &\text{May:} & F(x) &= \begin{cases} 
        0; & x < 2.7 \\ 
        -0.0062 \cdot x^3 + 0.114 \cdot x^2 - 0.4701 \cdot x + 0.5426; & 2.7 \leq x \leq 8.713 \\ 
        1; & x > 8.713 
    \end{cases}
\]

In accordance with Kolmogorov test according to (3) there was calculated the measure of deviation between the values of the empirical and theoretical distribution functions $D = 0.035$, the value $\lambda = 0.82$, $\lambda_{0.05} = 1.36$ (table 2). As far as $\lambda \leq \lambda_{0.05}$, the hypothesis that the distribution of water oxidability in January is defined by function (9), is adopted. For May, the value $D = 0.051$, the value $\lambda = 1.20$. As far as $\lambda \leq \lambda_{0.05}$, the hypothesis that the distribution of water oxidation in May is defined by the function (10), is adopted. The quality of approximation of empirical distributions on the intervals (7, 8) is very high, values of the determination coefficient $R^2$ are 0.9957 and 0.9935. Distribution functions (9, 10) accurately describe the curves of the empirical distributions of oxidability: the differences between the values of the theoretical and empirical distribution functions are extremely small (table 2, figure 4). On the basis of the obtained distribution functions of water oxidability, there have been estimated the probabilities of occurrence of any events significant for practice, taking into account the seasonal features of the process under research (table 3).

### Table 3. Estimation of the probability of exceeding the MPC\(^1\) of water oxidability.

| Months | Probability of oxidability not exceeding half of the MPC | Probability that MPC will not be exceeded | Probability that MPC will be exceeded not more than twice | Probability that MPC will be exceeded 3 times and more |
|--------|---------------------------------------------------------|------------------------------------------|----------------------------------------------------------|------------------------------------------------------|
|        | Range of oxidability values, mgO/dm\(^3\)               | under 2.5                                | under 5                                                 | from 5 to 10                                          | above 15                                             |
|        | Received with the use of gamma distribution             |                                          |                                                         |                                                      |                                                     |
| January| 0.756                                                   | 0.995                                   | 0                                                       | 0                                                    |
| May    | 0                                                       | 0.213                                   | 0.780                                                   | 0                                                    |
|        | Received with the use of polynomial (9, 10)             |                                          |                                                         |                                                      |                                                     |
| January| 0.742                                                   | 1.00\(^*\)                              | 0\(^*\)                                                 | 0\(^*\)                                              |
| May    | 0\(^*\)                                                 | 0.267                                   | 0.733                                                   | 0\(^*\)                                              |

\(^1\) MPC of water oxidability is 5 mgO/dm\(^3\)

\(^2\) The historical maximum of water oxidation in January is less than 15 mgO/dm\(^3\), so the appearance of large values is almost impossible.

\(^3\) The historical maximum of water oxidation in January is less than 5 mgO/dm\(^3\), so the appearance of large values - almost impossible.

\(^4\) The historical minimum of water oxidation for May is 2.7 mgO/dm\(^3\), so the appearance of lower values is almost impossible.

\(^5\) In accordance with the rule (10), the oxidability values must not exceed 5 mgO/dm\(^3\).
In a similar way, the distribution functions of water oxidability can be obtained for all other months.

4. Conclusion
In order to make it possible to take into account the seasonal features of the process under study, which form the dynamics of water oxidation, the distribution of water oxidation has been considered for each month separately. When empirical distribution does not coincide with any of the known theoretical distributions, it is suggested to approximate the series of cumulative relative frequencies by some continuous function which has all the properties of the distribution function.

Thus, computational experiments have shown that the distribution laws of water oxidation in January and May differ from the normal and logarithmic normal distributions. It is found that the distribution of water oxidation in January corresponds to the gamma distribution with parameters \( a = 4.814; b = 2.444 \), and can also be described by a third-degree polynomial \( F(x) = -0.0139x^3 + 0.0297x^2 + 0.4341x - 0.3113 \), but the acceptable region of values of argument \( x \) is limited to the interval from 0.695 to 3.73 mgO/dm\(^3\). The distribution of water oxidability in May corresponds to the gamma distribution with parameters \( a = 19.766; b = 3.228 \), and is also described by a third-degree polynomial \( F(x) = -0.0062x^3 + 0.114x^2 - 0.4701x + 0.5426 \), but at the same time the acceptable region of values of argument \( x \) is limited to the interval from 2.7 to 8.713 mgO/dm\(^3\).

On the basis of the obtained distribution functions of water oxidability, there have been estimated the probabilities of occurrence of any events significant for practice, taking into account the seasonal features of the process under research.

Gained laws of oxidability distribution offer the possibility to determine the probability that the values of the indicator will not exceed the specified value. Knowledge of the law of distribution of oxidability for each month helps to accurately estimate the probabilities of occurrence of any events significant for practice, which will allow evaluating the technical capabilities of water supply stations, taking into account the seasonal characteristics of the process under study, and developing solutions for water quality management.

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] Vozhdava M Yu, Kholova A R, Vagner E V, Kantor E A, Kantor L I, Trukhanova N V and Melnitsky I A 2018 The use of results of expanded monitoring research for the integrated assessment of drinking water according to indices of chemical harmlessness Gigiena i Sanitariya 97(2) 117-24
[2] Yalaletdinova A V, Enikeeva L V, Vozhdava M Yu and Kantor E A 2018 Statistical characteristics of relationships between turbidity and water flow rate caused by releases of the water reservoir Theoretical and applied ecology 1 33-42
[3] Yalaletdinova A V, Beloliptsev I I, Galimova Y O, Vozhdava M Yu and Kantor E A 2019 Probability analysis of water quality by turbidity IOP Conf. Ser.: Earth Environ. Sci. 315 062019
[4] Meyer Torsten, Lei Ying Duan and Wania Frank 2011 Transport of polycyclic aromatic hydrocarbons and pesticides during snowmelt within an urban watershed Water Research 45(3) 1147-56
[5] Vieira Rute, Fernandes Joao Nuno and Barbosa Ana Estela 2013 Evaluation, of the impacts of road runoff in a Mediterranean reservoir in Portugal Environmental Monitoring and Assessment 185(9) 7659-73
[6] Slipchenko A V, Kulsky L A and Matskevich E S 1990 The current state of the methods of oxidation of water impurities and the prospects for chlorination *Water Chemistry and Technology* **12**(4) 326-46

[7] Bigaliev A B, Sintyurina A V and Biyasheva Z M 2009 On the pathogenic effect of benzopyrene as an environmental pollutant (review) *Bulletin of KazNU. Ecological series* **1**(24) 14-21

[8] Karpukhin M M and Ladonin D V 2008 Influence of soil components on the absorption of heavy metals in conditions of technogenic pollution *Pochvovedenie* **11** 1388-98

[9] Kharabrin A V, Kharabrin S V, Kantor L I, Kantor E A and Klyavlin M S 2003 Comparison of water quality indicators of the Ufa River by turbidity, color, oxidizability and pH in sections of urban water intakes *Bashkir Chemical Journal* **10**(3) 82-3

[10] Kharabrin A V, Kharabrin S V, Kantor L I, Kantor E A and Klyavlin M S 2003 About the change in turbidity, color, permanganate oxidizability and pH of the water of the Ufa River *Bashkir Chemical Journal* **10**(3) 80-1

[11] Ayvazyan S A and Mkhitaryan V S 1998 *Applied Statistics And Introduction To Econometrics University Course Book* (Moscow: UNITY) p 1022

[12] Kremer N S 2009 *Probability Theory And Mathematical Statistics. University Ccourse Book* (Moscow: UNITY - DANA) p 551