Probability assessment of water turbidity in rivers based on optimizing the distribution function

A V Yalaletdinova¹, E A Kantor¹ and Y O Galimova²

¹ Department of Physics, Ufa State Petroleum Technological University, 1 Kosmonavtov St, Ufa, 450062, Russia
² National Library of Bashkortostan, Lenina st. 4, Ufa, 450077, Russia

E-mail: hawk22.89@mail.ru

Abstract. The method of least squares was used to optimize the theoretical water turbidity distribution function; as a result, the function was derived with all the properties of the distribution function (continuous and monotonically increasing function, values ranging from 0 to 1). It was demonstrated that the optimized polynomials can be used to estimate the probability of any water turbidity-related occurrence, but they cannot be used reliably for all time periods. Thus, the obtained polynomials are ineffective for studying the seasonal nature of changes in the turbidity distribution, which is not entirely convenient for modeling and entails the need to search for other mathematical models.

1. Introduction

Natural water is a multiphase system that contains various impurities. Therefore, to guarantee water treatment efficiency, especially during periods marked by increased risk of water pollution, thorough control of water quality is necessary as well as monitoring for key indicators such as water turbidity [1-3].

The analysis of the seasonal fluctuations in water turbidity values at the surface water intake will help to evaluate the capacity of water supply system, taking into account the seasonal changes affecting the water treatment process, and develop effective solutions for water quality management. Accurate prediction of the likelihood of adverse events associated with an increase in water turbidity can serve to improve and optimize the operation of water treatment facilities.

2. Research objectives and methodology

Previously, we collected and analysed daily statistical data on the turbidity of water at the surface water intake over the 17-year period from 1997 to 2014. [4, 5]. To account for the seasonal variations, the data for each month was considered separately. February was selected to best illustrate our findings, as it was typically the month preceding the flood [5].

Statistical analysis included the construction of an empirical distribution function for water turbidity, for the purposes of this paper satisfactorily described by the third-degree polynomial F(x). The resulting polynomial described the distribution of water turbidity and allowed us to predict the probability that the turbidity will not exceed the specified values of x [4, 5].

However, in order for the theoretical distribution function to match all the properties of the distribution function (continuous and monotonically increasing in the interval [0; 1]), the domain of the
function was limited to \( x \in [x_1; x_2] \subset [x_{\text{min}}; x_{\text{max}}] \), where \( x_{\text{min}}, x_{\text{max}} \) were the minimum and maximum values in the sample; \( x_1, x_2 \) were the values of the argument \( x \), when the function \( F(x) \) satisfied the set requirements. Polynomials obtained in Excel do not always meet these requirements (as evidenced by the above set conditions and the range of valid values of the \( x \) argument obtained for the functions used in our research).

Therefore, we propose to optimize the chosen theoretical distribution function \( F(x) \) and to determine the appropriate coefficients for the new optimized function \( F_o(x) \) where it will have all the properties of the distribution function, i.e. increasing monotonically and the domain of the function is from 0 to 1.

The coefficients were calculated by minimizing the sum of the squares of the differences, according to the method of least squares.

\[
S = \sum_{i=1}^{m} (F_n(x_i) - F(x_i))^2 \rightarrow \min
\]  

(1)

The following conditions must be met:

- \( F(0)=P(X<0)=0 \), since the water turbidity value cannot be negative;

\[
F(0) = 0
\]  

(2)

- the domain of the function is restricted to a range of values from 0 to 1 \([0;1]\]

\[
F(x_i) = a_0 + a_1x_i + a_2x_i^2 + a_3x_i^3 > 0
\]  

(3)

\[
F(x_i) = a_0 + a_1x_i + a_2x_i^2 + a_3x_i^3 < 1
\]  

(4)

- the value of the first derivative must be positive, then the function will be monotonically increasing;

\[
F'(x_i) = a_1 + 2a_2x_i + 3a_3x_i^2 > 0
\]  

(5)

The hypothesis that the empirical distribution matches the theoretical distributions is tested by the Kolmogorov cross-check test [6]. To do so, the discrepancy between the theoretical and empirical distribution is calculated

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

(6)

Then the value \( \lambda \) is calculated

\[
\lambda = D\sqrt{n}
\]  

(7)

where \( n \) is the output volume.

For the selected significance level \( \alpha \), the tabular value of \( \lambda_\alpha \) is determined [6]. If \( \lambda \leq \lambda_\alpha \), then the experimental data supports the hypothesis.

3. Results and discussion

Optimization was carried out for the presented earlier third-degree polynomial \( F(x) = 0.0623x^3 - 0.561x^2 +1.6833x-0.69 \), limited to \([0.485; 3.49]\), constructed to describe water turbidity distribution in February [5]. Based on the formulae (1-5) the coefficients were derived when the polynomial has all the properties of the distribution function (table 1).
Table 1. Coefficients for the optimized polynomial when the theoretical distributions of turbidity of water have all the properties of the distribution function.

| Month  | Coefficients of polynomials $F(x) = a_4 \cdot x^4 + a_3 \cdot x^3 + a_2 \cdot x^2 + a_1 \cdot x + a_0$ |
|--------|--------------------------------------------------------------------------------------------------|
| February | $a_4 = 0$ | $a_3 = -0.0069$ | $a_2 = -0.0505$ | $a_1 = 0.5469$ | $a_0 = 0$ |

The resulting polynomial is used to calculate the values of the optimized theoretical distribution function $F_d(x)$ (table 2).

Table 2. Characteristics of the theoretical distribution of water turbidity in February (after optimization): $N_0$ - interval number, $s_1$ - lower limit of the interval, $s^*$ - middle of the interval, $s_2$ - upper limit of the interval, $n_i$ - frequencies, $w_i$ - frequency (percentage), $F_d(s_2)$ - accumulated frequency, $F_d'(s_2)$ - values of the optimized distribution function, $F_d'(s_2)$ - first derivative (density).

| $N_0$ | $s_1$ | $s^*$ | $s_2$ | $F_d(s_2)$ | $F_d'(s_2)$ |
|-------|-------|-------|-------|------------|-------------|
| 1     | 0     | 0.25  | 0.5   | 0.018      | 0.260       | 0.520       |
| 2     | 0.5   | 0.75  | 1.0   | 0.527      | 0.490       | 0.460       |
| 3     | 1.0   | 1.25  | 1.5   | 0.768      | 0.683       | 0.388       |
| 4     | 1.5   | 1.75  | 2.0   | 0.934      | 0.837       | 0.307       |
| 5     | 2.0   | 2.25  | 2.5   | 0.986      | 0.944       | 0.215       |
| 6     | 2.5   | 2.75  | 3.0   | 0.998      | 1.000       | 0.113       |
| 7     | 3.0   | 3.25  | 3.5   | 1.000      | 1.000       | 0.000       |

Kolmogorov goodness of fit test was used to test the hypothesis regarding the function constructed to predict water turbidity distribution (table 3, table 4). According to (6), the discrepancy between the values of the empirical and theoretical distribution functions $D$ was calculated (table 3).

Table 3. Results of Kolmogorov goodness of fit test of the hypothesis (8).

| Table 3 | $s_2$ | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 |
|---------|-------|-----|-----|-----|-----|-----|-----|-----|
| $F_d(s_2)$ | 0.018 | 0.527 | 0.768 | 0.934 | 0.986 | 0.998 | 1.000 |
| $F_d'(s_2)$ | 0.260 | 0.490 | 0.683 | 0.837 | 0.944 | 1.000 | 1.000 |
| $|F_d(s_2) - F_d'(s_2)|$ | 0.242 | 0.038 | 0.084 | 0.097 | 0.042 | 0.002 | 0.000 |

The value of $\lambda$ is calculated by (7) (table 4). The tabular value for the significance level $\alpha = 0.05$ is $\lambda_{0.05} = 1.36$.

Table 4. Testing the hypothesis that the theoretical turbidity distribution function is defined by conditions (8), carrying out the Kolmogorov test (6, 7), $\lambda_{0.05} = 1.36$.

| Month  | $D$ | $\lambda$ | Conclusion |
|--------|-----|------------|------------|
| February | 0.242 | 5.40 | Since $\lambda > \lambda_0$, the hypothesis that the distribution of water turbidity in February is accurately described by function (8) is rejected |

Likewise, the optimization of the water turbidity distribution functions for all other months was carried out.

It was demonstrated that the optimized polynomials do not pass the Kolmogorov test, hence theoretical and empirical distributions of turbidity do not match. However, the empirical turbidity distribution function $F_d(x)$ for each month, separately, can be represented as the optimized $F_d(x)$ polynomials, for example, in February, May, August, October (figure 1).

Kolmogorov's test takes into account the maximum discrepancies between the empirical and theoretical distributions of the low turbidity values. However, when the values are large, the obtained polynomials do not diverge from the empirical distribution; therefore, in this case, the discrepancy
between the distributions is either minimal or non-existent, and the optimized theoretical distributions (polynomials) describe quite accurately the turbidity distribution obtained from the experimental data (figure 1). In applied water treatment, it is the large turbidity values that exceed the standard that are of greatest interest to us. Therefore, from a practical point of view, the obtained optimized polynomials can be used to estimate the probability of any water turbidity-related occurrence.

**Figure 1.** Values of empirical \( F_n(x) \) and theoretical (optimized) \( F_o(x) \) water turbidity distribution functions: \( a \) – February, \( b \) – May, \( c \) – August, \( d \) – October.

Based on the constructed water turbidity distribution functions, the probabilities of adverse events were estimated, taking into account the seasonal aspect of the changes in turbidity values (table 5).

| Month       | Water Turbidity Value, mg/dm³ | The probability of not exceeding the standard value | The probability of exceeding the standard by no more than 2 times the value | The probability of exceeding the standard by no more than 3 times the value | The probability of exceeding the standard by 5 times the value and/or more |
|-------------|-------------------------------|---------------------------------------------------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|
|             | less than 1.5                 | from 1.5 to 3                                   | from 1.5 to 4.5                                 | 7.5 and more                                    |
| January     | 0.616                         | 0.343                                            | 0.340                                          | 0.111                                          | 0.730                                          |
| February    | 0.683                         | 0.317                                            | 0.126                                          | 0.111                                          | 0.730                                          |
| March       | 0.630                         | 0.278                                            | 0.362                                          | 0.111                                          | 0.730                                          |
| April       | 0.021                         | 0.021                                            | 0.043                                          | 0.111                                          | 0.730                                          |
| May         | -4                            | -4                                               | 0.111                                          | 0.111                                          | 0.730                                          |
| June        | 0.393                         | 0.233                                            | 0.352                                          | 0.111                                          | 0.730                                          |
| July        | 0.443                         | 0.287                                            | 0.452                                          | 0.111                                          | 0.730                                          |
| August      | 0.604                         | 0.322                                            | 0.393                                          | 0.111                                          | 0.730                                          |
| September   | 0.811                         | 0.189                                            | 0.189                                          | 0.189                                          | 0.189                                          |
| October     | 0.568                         | 0.289                                            | 0.392                                          | 0.139                                          | 0.139                                          |
| November    | 0.262                         | 0.209                                            | 0.372                                          | 0.372                                          | 0.159                                          |
| December    | 0.440                         | 0.290                                            | 0.460                                          | 0.460                                          | 0.159                                          |

\(^3\) Water Turbidity Standard Value is 1.5 mg/dm³.

\(^4\) No Data

The obtained optimized polynomials are useful for calculating the risks of exceeding the specified values of the water turbidity, however, the functions are not equally applicable to all months. As a result, additional analytical tools and mathematical methods must to be used, which is not very convenient for modeling and forecasting.
4. Conclusion
Through the optimization of the water turbidity distribution function the appropriate coefficients were derived at which the approximating curve satisfies two important conditions: it is a monotonically increasing function and its values range from 0 to 1.

However, cross-check by the Kolmogorov Criterion revealed a discrepancy between the empirical and theoretical distributions. Thus, probability assessment demonstrated that the new optimized polynomials accurately predict the probability of the turbidity values not exceeding specified range. Nevertheless, for some months the optimized functions do not provide reliable data, which means that they are ineffective for use.

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] Zhigalova A V, Andreeva V A and Kantor E A 2016 Comparison of the turbidity of the water of the Ufa River in the sections and clean water tanks of urban water intakes 6th Int. Scientific and Practical Conf. of Young Scientists and Students: Experience of the Past - A Look into the Future vol 2 (Tula: Publishing House TulGU) 42-5
[2] 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
[3] 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
[4] Yalaletdinova A V, Enikeeva L V, Beloliptsev I I and Vozhdaeva M Yu 2018 Assessment of excess of permissible turbidity in the Ufa River for February 69th Scientific and Technical Conf. of students, graduate students and young scientists of USPTU vol 2 (Ufa: Publishing House USPTU) p 680
[5] Yalaletdinova A V, Beloliptsev I I, Galimova Y O, Vozhdaeva M Yu and Kantor E A 2019 Probability analysis of water quality by turbidity IOP Conf. Ser.: Earth Environ. Sci. 315 062019
[6] Kremer N S 2009 Probability Theory And Mathematical Statistics University Ccourse Book (Moscow: UNITY - DANA) p 551