

PAPER • OPEN ACCESS

## Probability analysis of water quality by turbidity

To cite this article: A V Yalaletdinova *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **315** 062019

View the [article online](#) for updates and enhancements.

## Probability analysis of water quality by turbidity

**A V Yalaletdinova<sup>1</sup>, I I Beloliptsev<sup>2</sup>, Y O Galimova<sup>3</sup>, M Yu Vozhdaeva<sup>1,4</sup> and E A Kantor<sup>1</sup>**

<sup>1</sup> Ufa State Petroleum Technological University, 1 Kosmonavtov st., Ufa, 450062, Russia

<sup>2</sup> Financial University under the Government of the Russian Federation, Ufa Branch, 69/1 Mustaja Karima st, Ufa, 450015, Russia

<sup>3</sup> National Library of Bashkortostan, 4 Lenina st., Ufa, 450077, Russia

<sup>4</sup> MUE Ufavodokanal, 157/2 Rossiyskaya st., Ufa, 450098, Russia

E-mail: hawk22.89@mail.ru

**Abstract.** The distribution of turbidity values in given sample is analyzed. The results demonstrated that the nature of the distribution of turbidity values during the year largely depends on the seasonal factor, hence the analysis of the distribution of turbidity is performed separately for each month. Order statistic (variation series) is computed and an empirical distribution function of turbidity values is derived. It is concluded, that the distribution of turbidity in given water sample differs from normal, log-normal and gamma distributions. However, it can be described with sufficient accuracy by a cubic polynomial function. The turbidity distribution hypothesis is tested by the Kolmogorov–Smirnov test. The water turbidity distribution function predicts the probability of exceeding the specified values of turbidity and enables numerical assessment of its likelihood.

### 1. Introduction

Most large cities rely on surface water sources for water supply, which allow for collection of a greater volumes of fresh water compared to groundwater. However, despite their great production capacity surface sources, which are most commonly rivers, are usually prone to significant seasonal fluctuations of all the major characteristics of water quality. Water turbidity value is one of the main parameters that determine the operation of water treatment facilities, strongly affected by the continuous seasonal changes in water quality.

To assess the potential risks associated with high levels of water turbidity, the seasonal nature of the turbidity value at the surface water intake site is studied and the probability is calculated for an occurrence when turbidity might exceed the specified standard.

### 2. Research objectives and methodology

Daily measurements of water turbidity values over the period from 1997-2014 at the surface water intake point (SWIP) are used in this study.

Statistical analysis was performed, order statistic, and relative frequencies histogram of turbidity were produced, the empirical distribution function was formed for the sample. The distribution law theory hypotheses were tested.



The original data series are used to produce the histogram of relative frequencies of turbidity. The range of turbidity values is divided into intervals of a certain length  $k$ , the limits for each interval ( $x_{mini}$ ,  $x_{maxi}$ ) are calculated, frequencies ( $n_i$ ) and relative frequencies ( $w_i = n_i / n$ ) of values falling into the  $i$ -th interval are determined.

Then, the turbidity distribution function  $F(x)$  was approximated based on a series of cumulative relative frequency values, which is assumed an empirical distribution function  $F_n(x)$ , derived from experimental data. The values of the cumulative relative frequency  $F_n(x)$  for all intervals, calculated as a cumulative total, lie in the interval from 0 to 1. Thus, the values of  $F_n(x)$  are interpreted as the empirical probability of the value of the random variable  $X$  not exceeding the value of the argument  $x$ :  $F(x) = P(X < x)$ . The approximation of the range of cumulative relative frequencies  $F_n(x)$  allows one to obtain a theoretical distribution function  $F^*(x)$ .

To test the hypotheses that the empirical distribution of water turbidity complies with normal and lognormal laws of probability distribution [1], Pearson's chi-squared test [2] was used, moreover, the hypothesis was tested that water turbidity follows a gamma distribution [1], which is described by the following density function

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

where  $a$  and  $b$  are gamma distribution variables;  $\Gamma(a)$  - Euler's gamma function.

Gamma distribution variables are derived from the following:

$$M(x) = \frac{a}{b}; \quad D(x) = \frac{a}{b^2} \quad (2)$$

Where  $M(x)$  and  $D(x)$  - expected value and variance, following the distribution law (2).

To estimate the values of the gamma distribution variables, the values of  $M(x)$  and  $D(x)$  are replaced by their sample estimates  $\bar{x}$  and  $s^2$ , thus, the variables are calculated as:

$$a = \frac{\bar{x}^2}{s^2}, b = \frac{\bar{x}}{s^2} \quad (3)$$

where  $\bar{x}$  - sample mean;  $s^2$  - sample variance

The hypothesis that empirical and theoretical distributions are consistent is tested by Kolmogorov–Smirnov test [2]. In order to test null hypothesis ( $H_0$ ), that the theoretical distribution function of a random sample is described by the equation obtained by approximation of the empirical distribution function, a deviation between the theoretical and empirical distribution is calculated:

$$D = \max |F_n(x) - \hat{F}(x)| \quad (4)$$

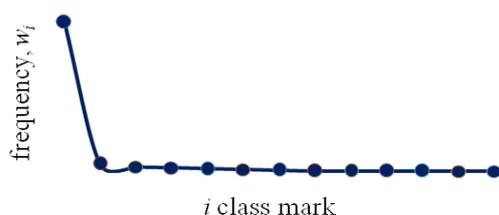
Then, value  $\lambda = Dn^{1/2}$  is found, where  $n$  is sample size. For the chosen statistical significance  $\alpha$  table value  $\lambda_{\alpha}$  is set. If  $\lambda \leq \lambda_{\alpha}$ , then the hypothesis  $H_0$  does not contradict the experimental data.

For the values of the argument, the random variable distribution function  $F^*(x)$  is continuous and increasing on the interval  $[0; 1]$  [2]. In addition, the approximation  $F^*(x)$  is produced based on the sample data, where range of values is limited to the minimum and maximum values  $[x_{min}; x_{max}]$ . Thus, the domain of the theoretical distribution function  $F^*(x)$ , when all properties of the distribution function will be maintained, is set under the following assumption:

$$x \in [x_1; x_2] \subset [x_{min}; x_{max}] \quad (5)$$

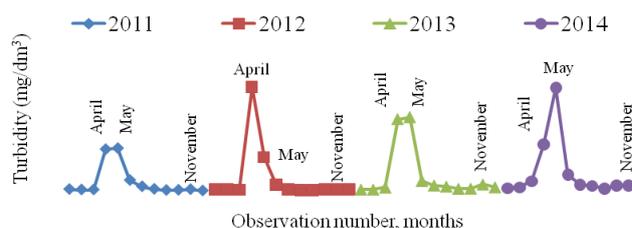
### 3. Results and discussion

The turbidity of the water increases dramatically in April-May every year, which is explained by the increased inflow of water in the river during these periods. However, the turbidity of water depends not only on the flow velocity, but also on many other factors, such as surface runoff from water collection point and the erosion of riverbed sediments [3-5]. Therefore, turbidity is considered a random variable. Based on the available sample, we constructed a range of relative frequencies and analyzed the distribution of turbidity over the entire observation period (figure 1).



**Figure 1.** Range of relative frequencies of water turbidity from 1997 to 2014.

Close to 90% of the observations fall into the first interval (up to 9 mg/dm<sup>3</sup>), the the remaining observations form the long sloping ‘tail’ of the distribution. The abnormally high values of turbidity are observed during seasonal spring floods, while in other months the turbidity is fairly stable and fluctuates within a certain interval. High turbidity values ( $M \geq 10$  mg/dm<sup>3</sup>) in September are extremely rare and its probability is approaching zero (figure 2).



**Figure 2.** Average monthly water turbidity values 2011-2014.

Therefore, it seems appropriate to investigate the distribution of turbidity for each month. The distribution of turbidity in February is considered in detail, since at this time the releases of water from the reservoir take place, associated with the need to regulate river flow in anticipation of spring floods [6].

The sample size  $n$  was 499. Over the entire observation period from 1997 to 2014 the lowest water turbidity reported in February was  $x_{min} = 0.5$  mg/dm<sup>3</sup>, and the highest -  $x_{max} = 3.2$  mg/dm<sup>3</sup>. Order statistic with an interval width of 0.5 is produced, the range of the intervals is determined, the relative frequencies  $w_i$  (figure 3) and the values of the empirical distribution function  $F_n(x)$  are calculated (table 1).

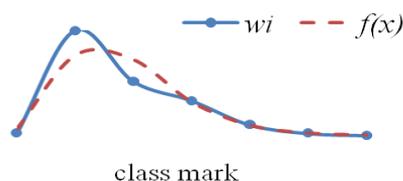
**Table 1.** Characteristics of the empirical, gamma and theoretical distributions of turbidity of water:  $N_0$  - interval number,  $s_1$  - lower limit of interval,  $s^*$  - class mark,  $s_2$  - upper limit of interval,  $n_i$  - frequency,  $w_i$  - frequency,  $F_n(s_2)$  - cumulative frequency,  $f(s^*)$  - the values of the gamma distribution density,  $F(s_2)$  - the values of the gamma distribution function,  $F'(s_2)$  - the values of the distribution function.

$N_0$	$s_1$	$s^*$	$s_2$	$n_i$	$w_i$	$F_n(s_2)$	$f(s^*)$	$F(s_2)$	$F'(s_2)$
1	0	0.25	0.5	7	0.014	0.014	0.0619	0.0572	0.019
2	0.5	0.75	1.0	248	0.497	0.511	0.7602	0.4144	0.495
3	1.0	1.25	1.5	128	0.257	0.768	0.7198	0.7679	0.783
4	1.5	1.75	2.0	83	0.166	0.934	0.3187	0.9319	0.931
5	2.0	2.25	2.5	26	0.052	0.986	0.0973	0.9836	0.985
6	2.5	2.75	3.0	6	0.012	0.998	0.0238	0.9966	0.993

No	$s_1$	$s^*$	$s_2$	$n_i$	$w_i$	$F_n(s_2)$	$f(s^*)$	$F(s_2)$	$F'(s_2)$
7	3.0	3.25	3.5	1	0.002	1.000	0.0050	0.9993	1.000

The hypotheses that the empirical distribution of water turbidity complies with normal and lognormal laws of probability distribution was tested with Pearson's chi-squared test [1] and was rejected.

The relative frequency range (figure 3) suggests the possibility that the turbidity of water in February has a gamma distribution [1].



**Figure 3.** The relative frequency range in February  $\omega_i$  and gamma distribution function curve  $f(x)$ .

When the values of the sample mean  $\bar{x} = 1.175$  and the sample variance  $s^2 = 0.257$  were put in (3) the distribution variables (1)  $a = 5.36$ ;  $b = 0.219$  and the values of the density function of the gamma distribution (table 1) were calculated. It is evident that the empirical data does not fully match the selected theoretical distribution (figure 3).

The hypothesis that turbidity of water follows gamma distribution was tested by Kolmogorov–Smirnov test [2]. The discrepancy between the values of the empirical and theoretical distribution functions was calculated for (4) (table 2).

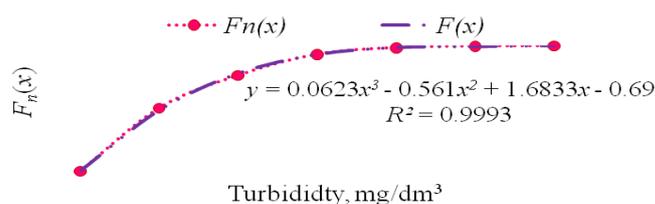
**Table 2.**Gamma distribution hypothesis Kolmogorov–Smirnov test results.

Distribution Hypothesis test results by (1)							
$s_2$	0.5	1.0	1.5	2.0	2.5	3.0	3.5
$F_n(X)$	0.014	0.511	0.768	0.934	0.986	0.998	1.000
$F'(X)$	0.057	0.414	0.768	0.932	0.984	0.997	0.999
$ F_n(s_2) - \hat{F}(s_2) $	0.043	0.097	0.000	0.002	0.002	0.001	0.001
Distribution Hypothesis test results by (8)							
$s_2$	0.5	1.0	1.5	2.0	2.5	3.0	3.5
$F_n(X)$	0.014	0.511	0.768	0.934	0.986	0.998	1.000
$F'(X)$	0.019	0.495	0.783	0.931	0.985	0.993	1.000
$ F_n(s_2) - \hat{F}(s_2) $	0.005	0.016	0.015	0.003	0.001	0.005	0.000

The value of  $D = 0.097$ ,  $\lambda = 2.159$ . The value for the statistical significance  $\alpha = 0.05$  is equal to  $\lambda_{0.05} = 1.36$ . Since  $\lambda > \lambda\alpha$ , the hypothesis that the turbidity distribution of water in February has a gamma distribution with parameters  $a = 5.36$ ;  $b = 0.219$  is rejected.

Since computational experiments show that the turbidity distribution of water does not correspond to any of the theoretical distribution laws discussed above [1, 2], an attempt was made to approximate the values of the empirical distribution function  $F_n(x)$  by a cubic polynomial (figure 4):

$$F'(x) = 0.0623x^3 - 0.561x^2 + 1.6833x - 0.69 \tag{6}$$



**Figure 4.** Empirical  $F_n(x)$  and theoretical  $F'(x)$  functions of turbidity distribution.

Approximation (6) is calculated based on sample where the minimum value  $x_{min} = 0.5 \text{ mg/dm}^3$ , and the maximum  $x_{max} = 3.2 \text{ mg/dm}^3$ , function (6) is positive, increases monotonically with values from 0 to 1 when  $x \in [0.485 ; 3.49]$ . Assuming (5), the range of permissible values of the argument  $x$  of function is the following (6):

$$x \in [0.485; 3.49] \subset [0.5; 3.2] = [0.485; 3.49] \tag{7}$$

So, water turbidity can assume any non-negative value. However, the observational data covers a fairly long period, when the value never exceeded  $3.2 \text{ mg/dm}^3$ . The possibility of exceeding this value exists, but the probability of such an event is extremely low. Therefore, in accordance with the principle of statistical significance, the probability that water turbidity does not exceed  $3.2 \text{ mg/dm}^3$  is equal to 1. Similarly, extremely low values of water turbidity (less than  $0.5 \text{ mg/dm}^3$ ) are almost impossible, and its probability is equal to 0. Based on these assumptions, the turbidity distribution function was set by the following:

$$\hat{F}(x) = \begin{cases} 0; & x < 0.485 \\ 0.0623 \cdot x^3 - 0.561 \cdot x^2 + 1.6833 \cdot x - 0.69; & 0.485 \leq x \leq 3.49 \\ 1; & x > 3.49 \end{cases} \tag{8}$$

To test the hypothesis Kolmogorov–Smirnov test was used. According to (4), the discrepancy between the values of the empirical and theoretical distribution functions  $D = 0.016$  was calculated, as well as the value  $\lambda = 0.367$ ,  $\lambda_{0.05} = 1.36$  (table 2). Since  $\lambda \leq \lambda_{\alpha}$ , the hypothesis that the turbidity distribution of water in February is described by function (8) is accepted.

The distribution function (8) fairly accurately describes the curve of the empirical distribution of turbidity: the differences between the values of the theoretical  $F'(x)$  and the empirical distribution functions  $F_n(x)$  are extremely small (table 2).

The distribution function of the argument  $x$  is equal to the probability that the value of the random variable  $X$  does not exceed the specified value. So, for example, the probability that the turbidity of water in February will be no more than  $1 \text{ mg/dm}^3$  is equal to:

$$P(X < 1) = \hat{F}(1) = 0.0623 \cdot 1^3 - 0.561 \cdot 1^2 + 1.6833 \cdot 1 - 0.69 = 0.495 \tag{9}$$

And the probability that the turbidity of water in February will exceed the values of  $1.5 \text{ mg/dm}^3$  by no more than 2 times:

$$P(1.5 < X < 3) = \hat{F}(3) - \hat{F}(1.5) = 0.993 - 0.783 = 0.210 \tag{10}$$

The quality of approximation of the empirical distribution within range (7) is quite high, with the determination coefficient of  $R^2 = 0.9993$ .

Based on this water turbidity distribution function, it is possible to estimate the probabilities of any significant changes (table 3).

**Table 3.** Estimated probability of exceeding specified turbidity values.

Turbidity values range, $\text{mg/dm}^3$			
up to 1.5	from 1.5 to 3	from 1.5 to 4.5	from 7.5 and higher
0.783	0.210	0.217	0 <sup>a</sup>

<sup>a</sup> The historical maximum water turbidity value in February is less than  $7.5 \text{ mg/dm}^3$ , therefore higher values are assumed practically impossible.

#### 4. Conclusion

The changes in turbidity values are largely due to seasonal fluctuations in the annual cycle. The seasonal nature of the turbidity must be taken into account and, therefore, the distribution of turbidity should be studied for each month separately. In case the empirical distribution is not consistent with any of the known theoretical distribution models, this study proposes to approximate the cumulative relative frequency series by some continuous function that will have all the properties of the distribution function.

Computational experiments have shown that the distribution of turbidity in February is different from the normal, log-normal and gamma distributions. It was demonstrated that the turbidity distribution in February can be described by a cubic polynomial  $F(x) = 0.0623x^3 - 0.561x^2 + 1.6833x - 0.69$ , but the range of acceptable values of the argument  $x$  is limited to 0.485 to 3.49 mg/dm<sup>3</sup>, because the distribution function of a random variable must adhere to certain properties.

The formula derived in this study describing distribution of water turbidity in February allows us to predict the probability of water turbidity exceeding a given value.

#### Acknowledgement

The work was performed within the framework of the State Task No. 5.12863.2018 / 8.9 “Development of a system for identification and quantitative analysis of environmental risks arising from the water supply of a large urban agglomeration”.

#### References

- [1] Ayvazyan S A and Mkhitarian V S 1998 *Applied Statistics And Introduction To Econometrics. University Course Book* (Moscow: UNITY) p 1022
- [2] Kremer N S 2009 *Probability Theory And Mathematical Statistics. University Course Book*. (Moscow: UNITY - DANA) p 551
- [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 for turbidity, color, oxidability and pH in the urban water intake sections *Bashkir Chemistry Journal* **10(3)** 82-3
- [4] Gao P, Pasternack G B, Bali K M and Wallender W W 2008 Estimating suspended sediment concentration using turbidity in an irrigation-dominated Southeastern California watershed *Journal of Irrigation and Drainage Engineering* **134(2)** 250–9
- [5] Stenberg L, Tuukkanen T, Finér L, Marttila H, Piirainen S, Kløve B and Koivusalo H 2015 Ditch erosion processes and sediment transport in a drained peatland forest *Ecological Engineering* **75** 421–33
- [6] Yalaletdinova A V, Enikeeva L V, Vozhdaeva 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