Probabilistic nature of water and environmental standards

O M Rosental\textsuperscript{1} and G A Sambursky\textsuperscript{2,3}

\textsuperscript{1}Water Problem Institute Russian Academy of Science, 3, ul. Gubkina, 119333, Moscow, RUSSIA
\textsuperscript{2}MIREA – Russian Technological University (M.V. Lomonosov Institute of Fine Chemical Technologies), 86, Vernadskogo Pr., 119571, Moscow, Russia
\textsuperscript{3}Technical Committee for standardization TC 343 «Water Quality» Rosstandart, 35, p. 2, Mosfilmovskaya st., 119330, Moscow, Russia

E-mail: gesamb@yandex.ru

Abstract. Determination of permissible concentrations of pollutants in relation to water bodies and water quality is the basis for further measures to form technological approaches to water treatment and wastewater treatment, to determine the standards of permissible exposure and technological indicators. The main standard, which is the basis for such calculations - the maximum permissible concentration (MPC) of pollutants or the approximate safe level of exposure (S), are threshold (deterministic) values. It is assumed that the establishment, for example, of the MPC for fisheries water bodies is based on the analysis of toxicological indicators. However, in fact, rationing is carried out on the basis of selective toxicological studies \cite{1}, which do not provide such determinism. The obtained data characterize only the influence of the selected levels of toxicant concentration on the studied organisms.

1. Introduction
The extension of the conclusions to the population, or, moreover, to the entire aquatic community at any concentration of the toxicant, creates representativeness errors that arise due to the discrepancy between the properties of the sample and the general population \cite{1} due to the variation of the controlled features and the difference in the distribution in the sample population from the distribution in the general population. Technological indicators of the best available technologies are established in approximately the same way – the basis for their deterministic purpose is not clear, and it is hardly necessary to talk about the environmental or water protection nature of such standards.

Substitution of the results obtained during the processing of random samples with deterministic values, both during rationing and during water quality control \cite{1,2}, is far from safe. As a result, for example, the water community may remain quite prosperous, despite the detected violation of quality standards, or oppressed when they are observed \cite{3,4}.

2. Methods
To form an acceptable rationing system, it is necessary to use statistical methods for planning experiments \cite{5,6}, since the suppression of organisms at each level of the concentration of a toxic substance, the number of such levels and the number of populations whose representatives are studied are random characteristics. The main integral indicator is the content of a toxic substance in the water. According to the terminology adopted in the analysis of variance, this is a factor or a predictive variable.
that is varied in order to determine its influence on the response describing the "factor effect" [7,8]. Accordingly, the rationing procedure should include a study of the effect of the factor by establishing statistical dependencies of intoxication of organisms on the concentration of an aqueous solution.

If the values \( n_j \) and \( \bar{l}_j \) are noticeably greater than one, then an approximation based on the normal distribution law can be used for further analysis. Then, for each concentration level, the boundaries of the confidence interval are equal (1):

\[
d_j \pm u_{1+\gamma}S_j(d_j)
\]

Here \( u \) is the quantile of the normal distribution, \( \gamma \) is the confidence probability,

\[
S_j(d_j) = \sqrt{\frac{D(\bar{d}_j)}{n_j\bar{l}_j(1-\bar{l}_j)}} - \text{estimation of the standard deviation of the number of dead organisms.}
\]

Therefore, the error of the estimated number of dead organisms at the \( j \)-th level at \( \gamma=0.95 \) (the value usually used in toxicological studies) is equal to (2):

\[
\Delta d_j = \pm 1.96\sqrt{n_j\bar{l}_j(1-\bar{l}_j)}.
\]

Note, however, that at the limit values of \( \bar{l}_j \), respectively 0 or 1, a normal approximation is impossible, and the calculation should be carried out according to the Klopper-Pearson equations [5,9]:

\[
\sum_{r=0}^{d_j} \frac{n_j^r}{r!(n_j-r)!}(\bar{l}_j)^r(1-\bar{l}_j)^{n_j-r} = 1 - \gamma_1,
\]

\[
\sum_{r=0}^{d_j-1} \frac{n_j^r}{r!(n_j-r)!}(\bar{l}_j)^r(1-\bar{l}_j)^{n_j-r} = \gamma_2,
\]

\[
\gamma_1 + \gamma_2 = 1 - \gamma.
\]

In particular, for \( d_j=0 \) (\( \bar{l}_j = 0 \)) or \( d_j = n_j \) (\( \bar{l}_j = 1 \)) the corresponding upper (index "B") and lower ("H") mortality limits are determined from the conditions (3):

\[
(1-\bar{l}_j)^{n_j} = 1 - \gamma \quad \text{and} \quad (\bar{l}_j)^{n_j} = 1 - \gamma. \quad (3)
\]

At the same time \( \Delta d_{j,H/B} = \bar{l}_j, \), it is necessary to check the condition that \( \pm d_j \) does not go beyond \((0, n_j)\), i.e. in relative units - (0,1). Otherwise, it is possible to use a truncated distribution. Then, for each calculated interval \( d_j \pm \Delta d_j \), the corresponding range of toxicant concentrations can be found:

\[
(C_{j,H}, C_{j,B})=C_j \pm \Delta C_j \quad \text{taking into account the dependence } d=f(C).
\]

3. Results and Discussion

3.1. Example 1. Processing of the results of a toxicological experiment for two populations of organisms

The results of monitoring the survival in a toxic environment of Pokemon snails (small-finned pond fish, Radix rubiginosa) and tropical guppy fish (the family of Peciliaceae, Poecilia reticulata) were used [9].

Intoxication of samples of organisms from 30 units of each population in aqueous solutions of iron (factor) was considered. The effect of the factor was the number of dead individuals in acute experiments (lethality). The results, interpolated in the range C= 100-150 mg/dm³, are shown in figure 1.
concentration

The number of dead organisms

Figure 1. Lethality of Pokemon snails (round markers) and guppy fish (square ones), as well as the corresponding trend lines (solid lines) and the MPC estimate, provided that the number of dead organisms \(d_0=3\) (dashed lines) is taken as the maximum permissible lethality.

Here, the concentration range of the toxicant (factor) equivalent to the lethality errors (factor effect) was determined using trend lines (figure 1), the analytical form of which is as follows (where the concentration dimension is mg/dm\(^3\)):

- for Pokemon snails \(d_j(C_j) \approx 0.002C^2 + 0.37C + 14.35\);
- for guppy fish \(d_j(C_j) \approx 0.001C^2 + 0.16C + 6.55\)

Table 1. Estimation of the error of the toxicological experiment for Pokemon snails (numerator) and guppies (denominator).

| \(J\) | \(C_j\) | \(d_j\) | \(\bar{l}_j\) | \(S^2_{d_j}\) | \(\pm \Delta_{d_j}\) | \(\pm \Delta_{d_j}/d_j\) | \([C_H + C_B]\) | \(\frac{C_B - C_H}{C_B}\) | \(\%\) |
|------|-------|-------|----------|------------|-------------|----------------|-------------|----------------|------|
| 1    | 100   | 1/1   | 0        | 0          | 0           | -             | 0            | 0              | 0    |
| 2    | 105   | 1/1   | 0.033    | 0.96       | 1.92        | 1.92          | 0+ 117       | 100            |      |
| 3    | 110   | 1/2   | 0.033    | 0.96       | 1.92        | 1.92          | 0+ 116       | 100            |      |
| 4    | 115   | 1/2   | 0.033    | 0.96       | 1.92        | 1.92          | 0+ 115       | 100            |      |
| 5    | 120   | 2/2   | 0.066    | 1.85       | 2.66        | 1.33          | 114+ 121     | 6              |      |
| 6    | 125   | 2/3   | 0.066    | 1.85       | 2.66        | 1.33          | 114+ 134     | 14             |      |
| 7    | 130   | 2/4   | 0.1      | 2.7        | 2.78        | 0.93          | 115+ 131     | 12             |      |
| 8    | 135   | 3/5   | 0.133    | 3.86       | 3.85        | 0.67          | 115+ 136     | 15             |      |
| 9    | 140   | 3/7   | 0.166    | 4.65       | 4.23        | 0.85          | 124+ 148     | 16             |      |
| 10   | 145   | 4/9   | 0.233    | 5.36       | 4.54        | 0.65          | 132+ 150     | 19             |      |
| 11   | 150   | 5/10  | 0.166    | 4.65       | 4.23        | 0.42          | 126+ 155     | ~140+ 180     |      |
| >10  | >0.33 | >7    | >5       | >0.5       | >0.5        | >20           | >20          | >20            |      |

Table 1: Estimation of the error of the toxicological experiment for Pokemon snails (numerator) and guppies (denominator).
It can be seen from the table that it is impossible to guarantee the absence of detail in the entire concentration range \( C_j > 0 \). Even if the estimated \( d_j = 1,2 = 0 \), the error \( \Delta d_j \) is so large that it was wrong to prevent the death of organisms due to the factor effect at all, unless, of course, the condition \( C = 0 \) is accepted.

On average, the width of the confidence intervals is 20-30 mg/dm3, and their overlap is observed for the two studied species of organisms. As a result, the apparent higher lethality of guppies compared to the lethality of Pokemon snails can also not be guaranteed. Let's assume, for example, that the "maximum inactive concentration" [1] is a 10% mortality rate \( (d_0 = 3) \). Then \( C_j = 140 \) mg/dm3 for Pokemon snails and \( C_j = 125 \) mg/dm3 for guppy fish (figure 1), and the boundaries of the confidence intervals, 121÷146 and 115÷131 (table 1), overlap in the range of 121÷131, so that the risk of mistakenly assigning a guppy as a more "sensitive test object " [1,5] is \( (131-121) / (146-115) = 0.32 = 32\% \).

It follows from table 1, in addition, that the error in estimating the MPC for the experimental results used with a low division price of the main scale (the difference in the concentration values corresponding to two adjacent scale marks does not exceed 5%) is relatively small: at \( d_0 = [1 ÷ 4] \) it does not exceed 20%. For example, if \( d_0 = 3 \) is selected above, we have \( \Delta_{\text{mpc}} \) is 12% for guppies and 17% for Pokemon snails. In practice, this error can increase dramatically. For example, if when performing "acute toxicological studies, 5-6 concentrations are selected that differ by 10 times" [1,9-11]. As a result, a real assessment of the MPC (or the risk) must be formed in the scheme of figure 2, according to which, in the case of a strict approach to compliance with the established requirements for water quality, compliance with the norm with the permissible level of mortality is accepted under the condition \( C < \text{MPC}_H \), and non-compliance-with \( C \geq \text{MPC}_B \). In other cases, it is sufficient to focus on the opposite boundary of the confidence interval, and then the correspondence is accepted under the condition \( C \leq \text{MPC}_B \), and the discrepancy at \( C > \text{MPC}_B \).

![Figure 2. Intervals of reliable rationing of water quality.](image)

### 3.2. Example 2. Scheme for evaluating the MPC as the maximum "inactive" concentration on the most sensitive test object

In accordance with [1,12], the link of the trophic chain that is most sensitive to the toxicant is defined as a limiting test object, according to which the MPC or the STANDARD are established. Therefore, it is necessary to compare the maximum permissible concentrations of pollutants for organisms of different populations.
The solution of this problem requires checking the statistical significance of the differences between the available data samples obtained in a toxicological experiment. In principle, such checks are also useful in the case of evaluating the effect of a factor at different levels of the latter. In this case, let the data of two samples be compared with the mathematical expectations of the concentration of the toxicant \( M(C_j) \) and \( M(C_{j+1}) \). If the large value is \( \{M(C_{j+1})>M(C_j)\} \), then they are indistinguishable under the condition of the following quantile inequality:

\[
u_0 = \frac{M(C_{j+1})-M(C_j)}{\sqrt{\sigma^2(C_{j+1})+\sigma^2(C_j)}} \leq u_{1-\alpha}, \tag{4}\]

where, according to the above found,

\[
\sigma_{C_j} = \frac{\Delta C_j}{1.96}, \tag{5}\]

\( \alpha \) is the significance level, i.e., the probability of rejecting the null hypothesis, if in fact it is true. Usually \( \alpha \sim (0.05+0.1) \), which corresponds to the values of the confidence probability \( (0.95\div0.9) \).

In the case of an unbiased (point) estimate, we have \( M(C_j)=C_j \) and \( M(C_{j+1})=C_{j+1} \), which simplifies the task of checking statistical distinctness, since then the data are not distinguishable under the condition

\[
u_0 = \frac{C_{j+1}-C_j}{\sqrt{\sigma^2(C_{j+1})+\sigma^2(C_j)}} \leq u_{1-\alpha}. \tag{6}\]

If \( u_0 > u_{1-\alpha} \), then the values are distinguishable, and the lower value \( C_j \) is chosen as the MPC.

Next, you can form hard and soft rules for making decisions about compliance with the standard for the concentration of the toxicant measured during monitoring: \( C_{\Pi} \): \( C_{\Pi} < C_{jH} \)-acceptance \( C_{\Pi} < C_{jB} \)-rejection. When \( C_{jH} \leq C_{\Pi} \leq C_{jB} \) a zone of uncertainty arises.

Similarly, you can check any two points, including to establish the most sensitive test organism.

### 3.3. Setting the maximum “inactive” concentration for one of the two test objects

The results of observing the survival rate in a toxic environment of Pokemon snails and guppy fish were used. Let us assume that 10% of the organisms of any of the two studied organisms of different populations, the 1st (snails) and the 2nd (guppies), are acceptable, so that \( l_1 = l_2 = 0.1 \). If the volume of samples of organisms \( n_1 = n_2 = 30 \), then \( d_1 = d_2 = 3 \). We also have \( \Delta d_1 = \Delta d_2 = 1.96 \sqrt{30 \cdot 0.1 \cdot 0.9} \approx \pm 3 \).

Hence, the lower bound of the confidence interval \( [d_1 - \Delta d_1, d_2 - \Delta d_2] \), i.e. it is at the maximum concentration, when there is still no death of organisms. If this is observed at \( C=100 \) mg/dm3 (table 1), and the permissible level of mortality of snails corresponds to the concentration of the toxicant \( C_1=140 \) mg/dm3 and guppies-125 mg/dm3, then the errors \( \Delta c_1 = 140-100 = 40 \) mg/dm3, \( \Delta c_2 = 125-100 = 25 \) mg/dm3. Therefore, \( \sigma_{c_1} = \frac{40}{1.96} \approx 20 \) mg/dm3, \( \sigma_{c_2} = \frac{25}{1.96} \approx 13 \) mg/dm3, \( \sigma^2_{c_1} = 400, \sigma^2_{c_2} = 169 \) and the quantile of the normal distribution \( u_0 = \frac{140-125}{\sqrt{400+169}} = 0.62 \).

With this value \( u_0 \) for the standard normal distribution function, we have \( 1-\alpha = 0.7324 \), i.e. \( \alpha = 0.27 \), which is significantly less than the accepted significance level \( \alpha = 0.1 \), at which \( 1-\alpha = 0.9 \) and, consequently, the quantile \( u_0 \), \( 1 = 1.28 \times u_0 = 0.62 \).

As can be seen, the difference in the calculated value of the criterion is less than the critical one and, therefore, is statistically insignificant (the significance level \( \alpha > 0.1 \)). Thus, according to the results of the toxicological experiment, a good coincidence of the parameters under consideration is observed. Therefore, any concentration can be chosen as the MPC. In practice, in such a situation, it is economically feasible to choose the maximum value: MPC = \( C_1 \). To solve environmental problems, we believe that the lowest concentrations in the calculated range should be chosen as the base for special
natural areas, for example, to establish requirements for discharge into the lake. Baikal. In this case, appropriate wastewater treatment will be technologically achievable. The maximum values in the MPC range can be a reference point for establishing technological indicators of the best available technologies, making the environmental measures carried out economically feasible.

4. Conclusions
The correct management of water use implies the rejection of the concept of absolute ecological safety of water management activities. Rationing, as well as water quality control, require a probabilistic (risk-oriented) approach, since the controlled toxicological or chemical-analytical characteristics are random values estimated by means of selective control. As a result, it is necessary to clarify the instructions of regulatory documents on the so-called "inactive concentrations" and on the MPC as a standard, in which "there are no consequences in the reservoir that reduce its fishing value" [4,13].

It is also necessary to form an approach that ensures a socially acceptable level of risk [7,14]. This means that the probability of undesirable or even dangerous consequences for the environment caused by water management activities is acceptable if the possible damage is so small that for the sake of the benefits received in the form of material and social benefits, the society is ready to “take a risk”. It is advisable to accept such a provision as a condition for evaluating the statistical error that characterizes the uncertainty of estimating the true values of water and environmental standards.

References
[1] Plitman S I, Tulakin A V and Sambursky G A 2016 Chemical substances Environment Health: reference guide ed N F Ismerov (Moscow: Technical literature publishing house) 384
[2] Bell S 2017 Urban Water Sustainability: Constructing Infrastructure for Cities and Nature 1-185
[3] Edzwald J K 2011 Water quality and treatment A handbook on drinking water. 6 ed. / ed.: J K Edzwald (N.Y.: McGraw-Hill) 1696
[4] Spellman F R 2014 Handbook for the treatment of natural and waste water Water supply and sewerage (SPb: TsOP Profession) 1312
[5] Rozental O M and Aleksandrovskaya L N 2021 Statistical Estimation of the Service Life of Water Treatment Systems Journal of Water Chemistry and Technology 3(1) 8-13
[6] Rozental O M and Aleksandrovskaya L N 2018 Assessment of the degree of compliance of water with regulatory requirements Water Resources 45(3) 379-87
[7] Samburskiy G A and Grodzenskiy S Y 2020 Approaches to risk assessment and selection of water treatment technologies to provide consumers with quality drinking water Amazonia Investiga 9(25) 33-43
[8] Makaeva V I, Akhmetzayanova G F and Makaeva A R 2012 About a cause of fish kill in the water objects of the republic of Tatarstan Veterinary doctor 6 34-6
[9] International Standard ISO 39511:2018 “Sequential sampling plans for inspection by variables for percent nonconforming (known standard deviation)” (ISO publ 2018-05)
[10] Recommendations for standardization P 50.1.040-2002. “Statistical methods. Design of experiments. Terms and definitions” (approved by the order of Gosstandart of Russia on 02 October, 2002 N 362)
[11] Mutzner L, Wani O, Mariane Y, Spuhler D, Beutler P and Maurer M 2017 The potential of knowing more – a review of data-driven urban water management Environmental Science & Technology 51(5) 2538-53
[12] Schaum C and Cornel P 2016 Sewage treatment of the future: health, water and resource protection Österreichische Wasser - und Abfallwirtschaft 68(3-4) 134-45
[13] Wurbs R A and Hoffpaur R J 2017 Environmental flow requirements in a water availability modeling system Sustainability of Water Quality and Ecology 9-10 9-21
[14] Samburskiy G A and Pestov S M 2017 Technological and organizational aspects of the processes of obtaining drinking water quality (Publishing Solutions) 184