PREDICTION OF BY-PRODUCT FORMATION DURING CHLORINE DIOXIDE DISINFECTION: FROM STATISTICAL TO HYBRID MODELS

Oleksandr Bondarchuk[1], *Arcady Shakhnovsky*[1], *Larysa Spasonova*[1], *Andriy Mokiienko*[2]
[1] Igor Sikorsky Kyiv Polytechnic Institute, Bondarchukau@gmail.com
[2] National University «Ostroh Academy», Mokienkoav56@gmail.com

DOI: https://doi.org/10.20535/2218-930012026359443

Safe drinking water supply remains one of the most critical humanitarian challenges facing Ukraine amid ongoing armed conflict. Destruction of water infrastructure, combined with deteriorating source water quality has created an urgent need for adaptive, scientifically grounded disinfection management tools. Chlorine dioxide offers significant advantages over conventional chlorination, including efficacy across a wide pH range, biofilm disruption capacity and absence of carcinogenic trihalomethanes. However, chlorine dioxide disinfection generates specific oxidation byproducts (primarily chlorites and chlorates) whose combined concentration is strictly limited under national and European regulations, creating a fundamental technological paradox between guaranteed pathogen inactivation and byproduct minimization. Presented study develops an integrated computer modelling framework for chlorine dioxide disinfection byproduct prediction and control. A hybrid model architecture was proposed as the foundation for an adaptive SCADA-integrated dosing system capable of real-time chlorine dioxide adjustment based on continuous water quality sensor inputs. A five-factor central composite rotatable design (pH, temperature, chlorine dioxide dose, contact time, total organic carbon) was implemented to build second-order polynomial response surface models for two target responses: the fraction of consumed chlorine dioxide as a measure of disinfection efficiency and chlorite concentration in treated water. Analysis of regression equations revealed the pH as the dominant factor controlling chlorite formation, with the free term of the chlorite formation model approaching the regulatory limit at average operating conditions, confirming that passive parameter maintenance cannot guarantee regulatory compliance. Multi-criteria optimization using the Harrington desirability function identified the optimal operating regime: pH 6.0, temperature 25 °C, chlorine dioxide dose 6.0 mg/dm^3, contact time 18.2 hours, TOC ≤ 10 mg/dm^3. A Pareto frontier analysis structured the full set of optimal trade-off solutions into three operational modes, providing water utilities with a flexible decision-making tool adaptable to current epidemiological priorities. Monte Carlo stochastic simulation quantified seasonal risk differentiation, demonstrating that pH reduction is a more effective control lever than dose reduction for maintaining chlorite compliance under summer source water conditions.

Keywords: *drinking water treatment, chlorine dioxide disinfection, chlorite formation, disinfection byproducts, adaptive dosing, response surface methodology, hybrid models*

Received: 20 March 2026 Revised: 09 April 2026 Accepted: 10 April 2026

1. Introduction

Ensuring the population’s access to safe and high-quality drinking water is one of the fundamental tasks of municipal infrastructure. Under conditions of full-scale military aggression against Ukraine, the water supply sector has experienced unprecedented stress: the destruction of pumping stations, filtration facilities, and distribution networks has led to a deterioration in both the quantitative and qualitative indicators of water supply. In particular, a persistent increase in turbidity and color of source water from riverine supplies has been observed, along with elevated concentrations of suspended solids and organic compounds due to the disruption of bank protection structures and uncontrolled discharges, as well as an increased risk of microbial contamination caused by damage to wastewater systems (Shumilova, 2023; Strokal et al., 2023; Hapich, Hennadii and Novitskyi, 2024). All these factors significantly complicate water treatment processes and necessitate a revision of established technological approaches.

Traditionally, drinking water disinfection in Ukraine has been carried out using chlorination, a technology with more than a century of application and a well-established performance profile. However, the interaction of chlorine with natural organic matter, including humic and fulvic acids, the concentrations of which have increased under wartime conditions, leads to the formation of trihalomethanes, such as chloroform, bromoform, and their derivatives, as well as haloacetic acids, which are recognized carcinogens (Mokiienko, 2026). According to the requirements of EU Directive 2020/2184, the total concentration of trihalomethanes must not exceed 100 mg/dm^3, while haloacetic acids must not exceed 60 mg/dm^3. In view of the gradual alignment of Ukrainian water legislation with EU standards within the framework of the Association Agreement, the application of chlorination in water bodies with elevated organic content is becoming increasingly problematic.

An alternative approach is the use of chlorine dioxide, an oxidant with a significantly improved safety profile in terms of the formation of chlorinated organic compounds. Chlorine dioxide retains its disinfecting activity over a wide pH range (5.0 to 10.0), effectively disrupts biofilms in distribution systems, and does not produce trihalomethanes in concentrations of hygienic concern (Sorlini and Collivignarelli, 2005; Mokiienko, 2025a; Mokiienko, 2025b). These advantages make chlorine dioxide particularly attractive in situations where source water quality is unstable and enriched with organic matter.

At the same time, the application of chlorine dioxide is associated with the formation of specific oxidation by-products, primarily chlorite ions and chlorate ions. For a long time, hygienic regulation focused mainly on chlorites, which was reflected in most early regulatory documents. However, the current WHO Guidelines for Drinking-water Quality (4th edition, 2022) and the updated requirements of EU Directive 2020/2184 introduce strict control of both compounds: the combined concentration of chlorine dioxide and chlorites must not exceed 0.7 mg/dm^3. Chlorates pose particular risks to pregnant women, infants, and individuals with glucose-6-phosphate dehydrogenase deficiency (Mokiienko, 2026). In addition, the decomposition of chlorine dioxide in alkaline media may result in the formation of trace amounts of chlorine and hypochlorite. However, these concentrations are typically negligible and do not lead to the formation of chlorinated organic by-products at levels characteristic of conventional chlorination (Mokiienko, 2025a).

Thus, a fundamental technological paradox arises: effective disinfection, especially under conditions of elevated microbial contamination, requires an increase in the chlorine dioxide dose and contact time, whereas the minimization of by-products, particularly chlorites, the formation of which increases nonlinearly with rising pH and temperature, dictates the opposite. This conflict represents a typical multicriteria optimization problem in which the objectives partially contradict each other.

Despite the considerable number of publications devoted to the chemistry of chlorine dioxide in water treatment, existing models for predicting by-product formation have significant limitations. First, the majority of studies consider chlorites or chlorates separately, without accounting for their simultaneous formation and regulatory constraints on their combined concentration (Sorlini and Gialdini, 2014; Ma et al., 2025). Second, widely used simplified linearized regression relationships (for example, Korn et al., 2002; Zanetti, 2008; Sorlini et al., 2016) do not adequately capture the nonlinear nature of chemical kinetics and interactions among factors, such as between pH and temperature during disproportionation. Third, kinetic and computational fluid dynamics models, although formally accurate, require large sets of input parameters and substantial computational resources, which limits their practical applicability in water utilities with constrained instrumental capacity. Machine learning approaches, including neural networks and support vector methods, demonstrate significant potential (McKelvey et al., 2025); however, their application requires large training datasets, which are difficult to obtain under conditions of unstable source water quality and wartime constraints. In particular, conducting classical full-scale experimental studies is complicated by damaged or partially functioning infrastructure. This substantially increases the role of computer modeling as a tool for generating operational technological recommendations based on limited but sufficiently reliable laboratory and operational data.

The aim of this study is to develop an integrated computer model of the water disinfection process using chlorine dioxide, which enables the prediction and minimization of an extended spectrum of by-products, including chlorites and chlorates, based on a combination of experimental-statistical modeling methods and optimization algorithms.

2. Materials and Methods

2.1 Analysis of Modeling Approaches

Effective control of the drinking water disinfection process using chlorine dioxide under conditions of unstable source water quality represents a complex multicriteria problem that cannot be fully resolved by any single class of mathematical models. In this study, a hierarchical modeling approach is proposed, comprising three levels of complexity and predictive accuracy. The first level consists of experimental statistical models, specifically response surface methodology, RSM (Suquet J. et al., 2021; El Houda et al., 2024), developed on the basis of designed experimental data and presented in this work as the primary outcome. The second level includes kinetic and hydrodynamic models, which enable a mechanistic interpretation of the obtained statistical relationships. The third level involves hybrid algorithms and machine learning methods that refine the predictions and provide a pathway toward the development of a full digital twin of the technological process. This section describes all three levels, as well as the tools of multicriteria optimization and stochastic analysis applied to the developed RSM models.

The most fundamental approach to describing the disinfection process is the construction of kinetic models that represent the chemical reactions of chlorine dioxide decomposition and the formation of its derivatives. In general, the kinetic model of chlorine dioxide consumption in an aqueous medium is described by a system of differential equations. In particular, for the equation describing the decay of reagent concentration in an ideally mixed system (a completely mixed reactor), the following expression is used (Song et al., 2023; Li et al., 2024):

$$\frac{dC_{ClO_2}}{dt} = -k_1 \cdot C_{ClO_2} \cdot C_{count} - k_2 \cdot C_{ClO_2} \cdot C_{Fe} - k_3 \cdot C_{ClO_2}^2, \quad (1)$$

s.t. C_{ClO_2} is chlorine dioxide concentration, mg/dm^3; C_{count} – content of organic impurities in water that are susceptible to oxidation by chlorine dioxide; C_{Fe} – content of iron compounds; k_i – reaction rate constants.

The first term represents the reaction of chlorine dioxide with organic matter, the second term corresponds to its interaction with reduced forms of iron, and the third term describes the self-decomposition of the reagent. All three rate constants depend on temperature according to the Arrhenius equation, which directly links the kinetic model to technological factors (temperature, pH) of the designed experiment.

It is important to emphasize the formation of chlorites and chlorates as coupled kinetic processes. The disproportionation of chlorine dioxide in an alkaline medium is described by the following reaction:

$$5ClO_2 + 4OH^- \rightarrow 4ClO_3^- + Cl^- + 2H_2O, \quad (2)$$

Mechanism (2) explains the dominant role of pH in chlorate formation, providing a mechanistic basis that is reflected in the coefficients of the RSM equations discussed below.

For real water treatment facilities with spatially non-uniform flow, it is also advisable to apply another class of models, namely distributed parameter models, which simultaneously account for advection, diffusion, and chemical transformation processes (Zhang et al., 2025; Bondarchuk et al., 2026):

$$\frac{\partial C_{ClO_2}}{\partial t} = D \cdot \frac{\partial^2 C_{ClO_2}}{\partial x^2} - v \cdot \frac{\partial C_{ClO_2}}{\partial x} - k \cdot C_{ClO_2}^2, \quad (3)$$

s.t. D is turbulent diffusion coefficient, m^2/sec; v is average flow velocity, m/sec; k is generalized reagent consumption (expenditure) constant.

Numerical solution of such equations using Computational Fluid Dynamics software makes it possible to determine the spatial concentration profile, that is, the variation of chlorine dioxide concentration C_{ClO_2} over time, in a contact reactor or pipeline, to identify dead zones with insufficient mixing of the reagent, and to optimize the location of dose injection as well as to estimate the residual concentration at the outlet of the contact chamber. It should be noted that RSM models provide a conservative lower-bound estimate of process performance for real reactors: the actual formation of by-products under full-scale conditions may differ from laboratory predictions due to hydrodynamic non-uniformity, which is explicitly accounted for in the CFD approach.

2.2 Design of the Experiment

The main dataset of this study was obtained through an active, designed multifactor experiment. A nonlinear central composite rotatable design with five factors was employed for conducting the experiments (Antony, 2023; Bondarchuk et al., 2025).

Five technological parameters that most significantly influence the disinfection process and the formation of by-products were selected as independent variables (factors): pH of the medium (X_1); temperature of the treated water (X_2), °C; total dose of chlorine dioxide (X_3), mg/dm^3; contact time (X_4), hours; and total organic carbon content (X_5), mg/dm^3.

The selection of these factors is justified by two considerations. On the one hand, they represent key operational parameters that water utilities can directly control or monitor in real time. On the other hand, each of them is included in the kinetic equations (1) and (3) as a variable determining reaction rates. In particular, pH directly controls the equilibrium of the disproportionation reaction (2), while total organic carbon content determines the first term in equation (1).

It should also be noted that the selected ranges of factor variation cover both typical operating conditions of water treatment plants in Ukraine under normal circumstances and extreme scenarios associated with deterioration of source water quality during wartime. In particular, the upper limit of total organic carbon, 30 mg/dm^3, corresponds to heavily polluted surface waters, while the wide temperature range, from 4 to 25 °C, encompasses seasonal variations from winter to summer conditions.

The response functions in this study are defined as follows: a) y_1 as the fraction of chlorine dioxide consumed for disinfection, a dimensionless quantity ranging from 0 to 1 (it should be noted that, in line with the study objectives, the residual volumes of chlorine dioxide in the purified water were determined; the y_1 function serves as an integral measure of the oxidative load and can be viewed as an indirect indicator of disinfection efficiency in terms of byproduct formation); b) y_2 as the concentration of chlorites in treated water, mg/dm^3, representing the primary regulated disinfection by-product.

Response Surface Methodology, RSM, was applied to approximate the response functions y_1 and y_2 within the factor space, enabling the construction and analysis of mathematical response models for optimization purposes (Harrington et al., 1992; Korn et al., 2002; Zanetti et al., 2008; Bondarchuk et al., 2025; Bondarchuk et al., 2026). The method is based on a second-order polynomial approximation:

$$y_u = \mathrm{a}_0 + \sum \mathrm{a}_i x_i + \sum \mathrm{a}_{ii} x_i^2 + \sum \mathrm{a}_{ij} x_i x_j, \quad (4)$$

s.t. y_u $(\mathrm{u} = 1, 2)$ are response functions (process responses), x_i $(\mathrm{i} = 1, .., 5)$ are independent variables (process factors), $\mathrm{a}_i, \mathrm{a}_{ij}$ $(\mathrm{i}, j = 1, .., 5)$ are the desired model coefficients.

The coefficient estimates were obtained using the least squares method. Statistical validation of the model included the calculation of coefficients of determination, testing the significance of coefficients using the Student t-test, and assessing model adequacy using the Fisher F-test (Bondarchuk et al., 2025).

The resulting response surface equations serve a dual purpose in this study. First, they act as analytical tools for understanding the influence of each factor on the target indicators. Second, they provide a basis for further computational analysis, including the construction of surface cross-sections, optimization, and stochastic modeling.

2.3 Multi-criteria optimization: desirability function and Pareto profile

The simultaneous minimization of several interdependent responses requires the application of multi-criteria optimization methods. In this study, two complementary approaches were employed.

Following Harrington’s generalized desirability function approach, each response y_i is transformed into a dimensionless partial desirability function d_i in the range [0, 1]. For a response to be minimized (by-product concentration), the transformation is defined as follows:

$$D = \left(\prod_{i=1}^{m} d_i^{w_i}\right)^{1/\Sigma w_i}, (5)$$

s.t m is number of responses.

The weighting coefficients wi reflect the priorities of a specific water utility: under an increased risk of microbial contamination, w1 (disinfection priority) is increased, whereas under strict regulatory control of water quality, w_2 (priority of chlorite minimization) is increased.

Optimization of the desirability function D with respect to the factors x_1, ..., x_5 yields a single optimum point that best satisfies the overall system of criteria.

To obtain a complete picture of the available optimal trade-offs between the responses, the Pareto front is also constructed as the set of solutions for which no response can be improved without deteriorating another. In the objective space (y_1, y_2), the Pareto front is identified numerically based on the RSM equations (4) using a uniform grid search in the factor space followed by filtering of non-dominated solutions. The practical value of the Pareto front for water utilities lies in providing decision makers with a comprehensive map of optimal operating regimes: the selection of a specific point on the front is determined by current priorities (degree of microbial risk vs risk of exceeding the maximum permissible concentration of chlorites), without the need to repeatedly solve the optimization problem.

2.6 Extension of the RSM approach capabilities

Despite the effectiveness of RSM for analysis and optimization within the designed factor space, this approach has fundamental limitations: a second-order polynomial approximation may fail to capture strongly nonlinear relationships, and a model built on 50 data points cannot be directly extrapolated to conditions that lie far beyond the studied range. These limitations can be addressed by machine learning and hybrid modeling methods (Hu et al., 2023; Ding et al., 2024), which form the basis for further development of the proposed approach. The network architecture for this case may include an input layer (5 neurons corresponding to the number of factors), several hidden layers with ReLU activation, and an output layer (neurons corresponding to the number of responses). A significant limitation for the application of ANN in the present study is the requirement for substantially larger training datasets (typically N >= 200 to 500 points); 50 central composite design points represent a minimum threshold and require mandatory use of regularization techniques (dropout, L_2) to prevent overfitting.

Support vector regression (SVR) constructs a hyperplane in the feature space with a maximum margin defined by an epsilon-insensitive tube, which ensures robustness to noise in the input data (Hu et al., 2021). In contrast to ANN, SVR with an RBF kernel (radial basis function) demonstrates acceptable performance even with small sample sizes (N approximately 30 to 80), making it particularly promising for direct application to the central composite design data of this study. An additional advantage is that SVR provides not only point predictions but also an estimate of the uncertainty interval, which is critical for assessing the risk of exceeding the maximum permissible concentration of chlorites.

A promising direction is the development of a hybrid (grey-box) model that combines a mechanistic basis (RSM equations or a kinetic model) with a machine learning component that describes systematic residuals:

$$\hat{y}_{hybr}(x) = \hat{y}_{RSM}(x) + \hat{\varepsilon}_{ML}(x), \text{(6)}$$

s.t $\hat{y}_{RSM}(x)$ is "traditional" (physically interpreted, limited by quadratic approximation) part of the model, $\hat{\varepsilon}_{ML}(x)$ is a residual term trained using SVR or ANN methods on nonlinear effects beyond the limits of quadratic approximation.

This approach, referred to in the literature as “physics informed machine learning” or “grey box modeling” (Hu et al., 2021; Moeini and Abokifa, 2025; Di Bella et al., 2026), combines the advantages of both classes of models: the physical interpretability and extrapolation robustness of RSM with the flexibility of machine learning in approximating complex nonlinearities. The RSM models developed in this study represent a natural first level of such a hybrid architecture and can be further enhanced with a machine learning component as operational data from water treatment plants are accumulated.

Deterministic RSM models provide a point prediction of the response at fixed factor values. However, real operating conditions of water treatment plants are characterized by continuous stochastic fluctuations in raw water quality parameters, such as temperature, pH, and organic content, which necessitates a probabilistic approach to risk assessment.

Finally, the Monte Carlo method (de Brito Cruz et al., 2024) enables the use of RSM equations as a generator of probabilistic predictions. At the first stage, probability distributions are assigned to the fluctuating factors, for example:

- $X_2 \sim N(\mu_T, \sigma_T^2)$ for the temperature factor (where N(..) is normal distribution with seasonal parameters);
- $X_1 \sim U(a_{pH}, b_{pH})$ for the pH factor (where U(..) is uniform distribution in the typical operating range);
- $X_5 \sim LogN(\mu_{TOC}, \sigma_{TOC}^2)$ for the TOC factor (where LogN(..) – lognormal distribution, typical for natural waters).

In the second step, an array $x^{(i)}$ of NN = 10,000 random vectors is generated according to the given distributions. In the third step, the RSM model prediction is calculated for each vector, for example, $\hat{y}_2^{(i)} = \hat{y}_2(x^{(i)})$. Finally, the probability of exceeding the standard is estimated:

$$P\left(\hat{y}_2 > \mathrm{MPC}_{y_2}\right) = \frac{1}{N} \cdot \sum_{i=1}^{n} 1\left[\hat{y}_2^{(i)} > 0.2\right], \text{(7)}$$

s.t. $P(\ldots)$ is mathematical probability, MPC_{y_2} is Maximum Permissible Concentration (MPC) for y_2, $1[\ldots]$ is indicator function.

This approach makes it possible to address practically important questions: what is the probability of exceeding the MPC of chlorites under winter temperature conditions compared to summer conditions; how safe the current dosing regime is under deterioration of raw water quality (for example, TOC spikes after floods or emergency events); at what value of chlorine dioxide dose the probability of exceeding the MPC does not exceed a specified risk threshold (for example, 5%), and so forth. The answers to these questions, obtained on the basis of RSM equations as a computational model (a digital twin), have direct practical value for automated reagent dosing control systems at water utilities.

3. Results and Discussion

3.1 Implementation of the experimental design and obtained arrays of technological responses

In accordance with the central composite rotatable design for five factors, a total of 50 experiments were conducted. A fragment of the design matrix in natural values of the factors (X_1 to X_5), together with the measured response values (y_1, y_2), is presented in Table 1. A characteristic feature of the obtained dataset is the wide range of response variation: y_1 changes from 0 (no activation of the reagent under conditions of minimum temperature and minimal contact time in the absence of organic matter) to values close to 0.92 (almost complete consumption of chlorine dioxide at maximum temperature, dose, and TOC content). The response y_2 varies from 0 to values close to 0.61 mg/dm^3, which in a number of cases exceeds the maximum permissible concentration of chlorites, thereby confirming the relevance of the dosing optimization problem.

Table 1. *Fragment of the experimental plan implementation matrix (natural values of factors)*

X_1 (pH)	**X_2 (°C)**	**X_3 (mg/dm^3)**	**X_4 (hours)**	**X_5 (mg/dm^3)**	**y_1**	**y_2 (mg/dm^3)**
6	25	6	20	30	0.00	0.25
9	4	1	20	30	0.00	0.00
6	4	6	20	10	0.68	0.28
7.5	39.47	3.5	10.5	20	0.39	0.38
7.5	14.5	3.5	10.5	20	0.28	0.27
9	25	6	20	10	0.92	0.61
9	25	1	1	10	0.42	0.20
6	25	6	1	10	0.25	0.00

Based on the responses, the matrix least squares method applied to the full matrix of 50 experiments, and regression equations for the responses were obtained.

It should be noted that the resulting polynomial equations are second order models and, like any quadratic approximation, have an inherent limitation: they accurately reproduce the system behavior within the studied factor space, but may systematically deviate from actual values when extrapolated beyond the design domain and in the presence of higher order nonlinearities not captured by the quadratic polynomial. To address these limitations, a hybrid grey box model (6) is proposed, combining a mechanistic RSM basis with a machine learning component trained on systematic residuals. Within this study, residual vectors (approximation errors) were calculated for each of the 50 design points and for each technological response. Testing the statistical hypothesis of no autocorrelation in the set of residual values using the Durbin Watson criterion showed that the residuals are randomly distributed and do not exhibit a systematic structure. Therefore, the obtained RSM models are adequate, and further machine learning correction is not justified, as it would not provide a significant improvement in accuracy.

3.2 Analysis and Interpretation of Regression Models

In the equation for y_1, all quadratic terms have negative signs. This indicates the presence of a pronounced maximum of the response surface y_1 within the investigated factor space. In contrast, in the equation for y_2, the quadratic terms for x_3 (+0.02) and x_4 (+0.02) have positive signs, whereas those for x_1, x_2, and x_5 are negative, forming a saddle-shaped surface for y_2 with a local minimum in the region of moderate doses and contact times, and an increase in chlorite concentration at extreme values of both factors.

3.2.1 Model y_1 (Disinfection Efficiency)

The intercept $a_0 = 0.54$ represents the average fraction of chlorine dioxide consumed at the center of the factor space (pH = 7.5; t = 14.5 °C; dose = 3.5 mg/dm^3; t = 10.5 hours; TOC = 20 mg/dm^3), corresponding to typical operating conditions of a water treatment plant.

Analysis of the linear coefficients reveals a noteworthy result: factor X_1 (pH) has a negative linear coefficient ($a_1 = -0.049$), indicating that an increase in pH within the studied range slightly reduces the effective useful consumption of chlorine dioxide for disinfection. A possible technological interpretation is that at higher pH values, a larger fraction of the reagent is consumed in disproportionation reactions, that is, in the formation of chlorites and chlorates, rather than in pathogen inactivation. At the same time, X_5 (total organic carbon) has the largest linear coefficient in absolute value among the factors except for x_1 ($a_5 = +0.05$), meaning that higher organic content increases the overall consumption of the reagent; however, it does not necessarily enhance microbiological disinfection efficiency, since chlorine dioxide is consumed in competing reactions with natural organic matter.

Among the interaction effects, the term x_2x_4 with a coefficient of +0.072 stands out as the largest in absolute value among all cross terms. This indicates that the effect of contact time (x_4) is significantly amplified at elevated temperature (x_2). In particular, at t = 25 °C, increasing contact time leads to a much greater increase in y_1 than at t = 4 °C. In practical terms, this suggests that at low winter temperatures, compensating for the reduced rate of oxidation reactions solely by increasing contact time is inefficient; an increase in reagent dose is required.

All quadratic coefficients are negative, with the largest absolute values being $a_{22} = -0.086$ for temperature and $a_{44} = -0.082$ for contact time. This indicates that the response surface y_1 exhibits a pronounced dome-shaped maximum in the x_2 dimension: excessive increases in temperature or contact time beyond the optimal level leads to a decrease in y_1, likely due to the acceleration of side reactions and decomposition processes.

3.2.2 Model y_2 (Chlorite Concentration)

The key linear effect is $a_1 = -0.028$, which is the largest linear coefficient in absolute value in the equation for y_2. The negative sign indicates that a decrease in pH leads to a reduction in chlorite concentration. At lower pH values, the equilibrium of chlorine dioxide disproportionation shifts to the left, and the formation of chlorites via reaction (2) is slowed down.

The most significant interaction effect in the y_2 model is the term x_1x_5 with a coefficient of -0.064. This is the largest interaction coefficient in absolute value across both equations. The negative sign indicates a synergistic reduction of y_2 under conditions of simultaneously low pH and high total organic carbon, compared to what would be expected based on linear effects alone. Under such conditions, chlorine dioxide predominantly oxidizes the organic matrix without significant formation of chlorites through the disproportionation mechanism.

3.3 Analysis of Response Surfaces and Determination of the Safe Dosing Region

Since full visualization of the response space for a five-factor system is not feasible, as the response surface represents a five-dimensional function and any cross-section is only a partial projection, graphical analysis in this study was performed using the method of cross-sections. In this approach, two factors were varied simultaneously, while the remaining three were fixed at specified levels. A critically important conclusion is that the shape of the cross-section fundamentally depends on the levels of the fixed factors, as illustrated in Figure 1. In particular, the response surface constructed in the x_1-x_2 space for the variable y_1 differs significantly depending on the levels at which the remaining factors are fixed. In the case of fig. 1, b (short contact duration, $x_3 = 0, x_4 = -1,$ and $x_5 = 0$), a surface maximum is observed in the zone $x_1 \approx -0.5$, $x_2 \approx +0.5$ (moderately acidic pH, elevated temperature). In the case of fig. 1, a (long contact duration, $x_3 = 0, x_4 = 1,$ and $x_5 = 0$), the indicated maximum shifts and "flattens" – long contact compensates for the influence of pH and temperature, leveling out the difference between the conditions.

The safe dosing region is determined graphically and analytically as the intersection of the domains satisfying multiple conditions, for example, simultaneous fulfillment of $\hat{y}_1 \geq 0.5$ condition, effective disinfection defined as at least 50 percent of chlorine dioxide consumed, and $\hat{y}_2 \leq 0.2$ condition, compliance with the maximum permissible concentration for chlorites.

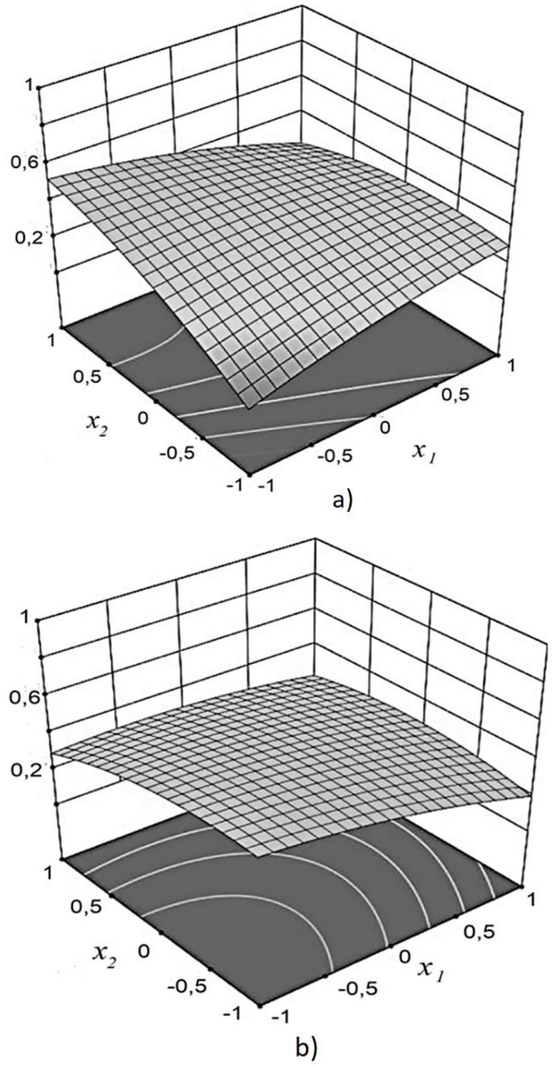

Fig. 1. *Response surface plots: a case study of disinfection efficiency*

Superposition of the contour plots of both responses shows that this region is narrow and highly sensitive to the values of the fixed factors. Under unfavorable combinations of background conditions, for example, pH ≈ 9, TOC ≈ 30 mg/dm^3, the region either becomes significantly reduced or disappears entirely, making it impossible to satisfy both regulatory requirements simultaneously for any combination of dose and contact time.

3.4 Numerical Multicriteria Optimization

The conflicting nature of the objective functions, with y_1 to be maximized and y_2 to be minimized, significantly complicates the quantitative identification of a point that simultaneously optimizes both responses. To obtain a compromise solution, the Harrington approach was applied (Zalenska, 2023): each response was transformed into a dimensionless partial desirability function di in the range from 0 to 1, and the overall process quality criterion was maximized using an evolutionary algorithm over the full factor space for the given case. The transformation of responses into desirability functions was performed with consideration of regulatory constraints. In particular, for y_2, the threshold of desirability was set at the level of the MPC = 0.20 mg/dm^3 (d_2=0 with $\hat{y}_2 \geq 0.2$ mg/dm^3, and d_2=1 with $\hat{y}_2 = 0$.

The results of numerical optimization gave the following optimal values of factors in coded form: $x_1 = -1, x_2 = 1, x_3 = 1, x_4 = 0{,}81$, $x_5 = -1$, or in natural units $X_1 = 6$, $X_2 = 25$ °C, X_3 =6 mg/dm^3, $X_4 = 18{,}2$ hours, $X_5 = 10$ mg/dm^3.

The physicochemical interpretation of the optimum found is consistent with the mechanism of the reactions studied. The minimum pH (pH=6.0) maximally suppresses the disproportionation of chlorine dioxide, reducing y_2. The maximum temperature (25 °C) accelerates the disinfecting reactions, increasing y_1. The maximum dose (6.06 mg/dm^3) in combination with an acidic environment provides a powerful oxidative potential with relatively moderate formation of chlorites. The duration of 18.2 hours (close to the upper limit) aims to ensure exhaustive disinfection.

3.5 Pareto front: set of optimal compromise solutions

To achieve a systematic understanding of the trade-off structure between y_1, efficiency, and y_2, safety, a Pareto front was constructed as the set of non-dominated solutions in the objective space. For this purpose, based on the equations in the factor space $x_i \in [-1; +1]$, a uniform grid with a step of 0.25 for each factor was generated. For each point, the pair of predicted responses $(\hat{y}_1, \hat{y}_2)$ was calculated. The solution x^A dominates over x^B if in the $\hat{y}_1^A \geq \hat{y}_1^B$; $\hat{y}_2^A \leq \hat{y}_2^B$ system although one of the inequalities is a strict inequality. After selection according to this criterion, the Pareto front containing N_p non-dominated solutions was obtained.

Analysis of the Pareto front revealed three functionally distinct regimes:

- Mode A – maximum disinfection ($\hat{y}_1 > 0.7,\ \hat{y}_2 > 0.2\ mg/dm^3$). Mode A is characterized by the following parameters: maximum temperature within the range, maximum oxidant dose, and moderate or elevated pH value. In this mode, the regulatory limit for y_2 is not satisfied, but disinfection efficiency is maximal. This regime may be applied under conditions of elevated microbiological risk, when the risk of infection outweighs the risk of chemical contamination.
- Mode B – minimal chlorites ($\hat{y}_1 < 0.4$, $\hat{y}_2 < 0.1\ mg/dm^3$). Mode B corresponds to minimum pH, reduced oxidant dose, and shortened contact time. In this regime, the regulatory requirement for y_2 is satisfied with a significant margin, but disinfection efficiency is insufficient for waters with high microbial load.
- Mode C represents the corridor of compromise optimum ($0.5 \leq \hat{y}_1 \leq 0.7$, $0.15 \leq \hat{y}_2 \leq 0.20\ mg/dm^3$). The optimal

solution of the multicriteria optimization problem identified above lies within this corridor. It defines the operational zone of normal plant performance, where both regulatory requirements are satisfied at an acceptable level of disinfection efficiency.

The principal practical value of the Pareto front lies in its ability to transform the selection of a dosing regime from a single-point decision into a flexible management tool. Movement along the front from Mode B to Mode A under the threat of a microbiological outbreak, and in the opposite direction under stricter regulatory control of water quality, can be implemented automatically by a SCADA system based on signals from source water quality analyzers.

3.6 Stochastic analysis of models for seasonal risk assessment

Under real operating conditions at a water treatment plant, the quality parameters of source water continuously fluctuate. To quantitatively assess the probability of exceeding the maximum permissible concentration of chlorites under stochastic variations of conditions, the Monte Carlo method, that is, statistical simulation, was applied.

Two seasonal scenarios were analyzed with controlled factors fixed at their optimal levels, $x_3^* = +1$ and $x_4^* = +0.8$, and with stochastic variations in uncontrolled conditions, in particular:

- Winter scenario: $X_1 \sim U(7.0,7.8)$; $X_2 \sim N(7, 2^2)$; $X_5 \sim LogN(15, 3^2)$.
- Summer scenario: $X_1 \sim U(7.4,8.2)$; $X_2 \sim N(21, 2^2)$; $X_5 \sim LogN(24, 4^2)$ (increased TOC due to blooms in water bodies).

The uniform distribution of X_1 factor reflects the fact that pH within the operational range varies relatively evenly, and the operator maintains it within specified limits. The lognormal distribution of total organic carbon is selected in accordance with its natural distribution in surface waters.

For each scenario, NN = 10,000 random vectors of factor values were generated. The probability of exceeding the maximum permissible concentration was estimated using equation (7).

Statistical simulations conducted according to the described procedure revealed a pronounced difference between seasons. In the winter scenario, lower temperature slows down the disproportionation reaction of chlorine dioxide, while lower organic content reduces the competitive consumption of the reagent. In the summer scenario, the simultaneous increase in temperature, by 14 °C relative to the center of the factor space, pH, with an upward shift of the distribution by 0.4 units, and total organic carbon, by 9 mg/dm^3, leads to a significant rightward shift in the distribution of the predicted response $\hat{y}_2$.

The practical conclusion from the results of the stochastic analysis is as follows: the optimal dosing regimen proposed above, aimed at minimum TOC conditions, is adequate for the winter season, but most likely requires correction (by reducing the chlorine dioxide dose or lowering the operating pH to compensate for the increased risk of exceeding the MPC) for summer conditions.

4. Conclusions

The present study is devoted to the development of an integrated system of experimental statistical mathematical models of the drinking water disinfection process using chlorine dioxide, aimed at predicting and minimizing disinfection by products under conditions of unstable raw water quality. The analysis of the feasibility of hybrid models has shown that the experimental statistical models obtained in this case are adequate, and further refinement using machine learning methods is not justified, as it does not provide a significant improvement in accuracy.

Multi-criteria optimization based on the Harrington desirability function made it possible to propose an optimal technological regime of the process that ensures effective disinfection while maintaining chlorite concentrations below the regulatory limit. The constructed Pareto front structured the set of optimal trade off solutions into three functional regimes: maximum disinfection, minimum by products, and an operational compromise corridor, thereby providing water utilities with a tool for flexible selection of operating conditions depending on current epidemiological priorities and regulatory requirements.

Stochastic analysis of the models using the Monte Carlo method quantitatively confirmed the seasonal differentiation of risk: under a summer raw water quality scenario, lowering the operating pH is a more effective measure for maintaining chlorite concentrations within the maximum permissible concentration than a proportional reduction in reagent dose. The obtained regression equations are suitable for direct implementation in SCADA systems of water treatment plants to enable proactive real time adjustment of chlorine dioxide dosing, which is critically important under conditions of martial law and increased variability of raw water quality in Ukraine.

Acknowledgments

The present study was carried out within the framework of the research projects “Technological Principles for Minimizing Chlorite Content in Drinking Water after Disinfection with Chlorine Dioxide” (state registration number: 0125U001784) and “Computer Modeling and Optimization of Sustainable Technological Schemes in Water Management” (state registration number: 0124U002127).

References

1. Antony, J. **2023**. *Design of Experiments for Engineers and Scientists*, 3rd ed.; Elsevier.
2. Bondarchuk, O. V.; Shakhnovsky, A. M.; Spasonova, L. M.; Mokiienko, A. V. **2025**. Water Disinfection with Chlorine Dioxide: Experimental-Statistical Modeling. In *Water Supply and Wastewater Disposal: Design, Construction, Operation, Monitoring*; Malovanyi, M. S., Ed.; NGO “MNG”: Kyiv, Ukraine; pp 121–136.
3. Bondarchuk, O.; Spasonova, L.; Shakhnovsky, A.; Kvitka, O. **2026**. Study of Water Treatment Processes with Chlorine Dioxide: 1. Experimental Design. *Herald Khmelnytskyi Natl. Univ., Tech. Sci.* 363 (2), 237–243. https://doi.org/10.31891/2307-5732-2026-363-32
4. de Brito Cruz, D.; Schmidt, P. J.; Emelko, M. B. **2024**. Drinking Water QMRA and Decision-Making: Sensitivity of Risk to Common Independence Assumptions about Model Inputs. *Water Res*. 259, 121877. https://doi.org/10.1016/j.watres.2024.121877
5. Di Bella, A.; Raissi, M.; et al. **2026**. Physics-Informed Neural Networks in Water and Wastewater Systems: A Critical Review. *Water Res*. 293, 125449. https://doi.org/10.1016/j.watres.2026.125449
6. Ding, Y.; Sun, Q.; et al. **2024**. Application of Artificial Intelligence in (Waste)Water Disinfection: Emphasizing the Regulation of Disinfection By-Products Formation and Residues Prediction. *Water Res*. 253, 121267. https://doi.org/10.1016/j.watres.2024.121267
7. El Houda Mellal, N.; Tahar, W.; et al. **2024**. Prediction of Purified Water Quality in Industrial Hydrocarbon Wastewater Treatment Using an Artificial Neural Network and Response Surface Methodology. J.

Water Process Eng. 58, 104757. https://doi.org/10.1016/j.jwpe.2023.104757

8. Hapich, H.; Novitskyi, R.; et al. **2024**. Water Security Consequences of the Russia–Ukraine War and the Post-War Outlook. *Water Secur*. 21, 100167. https://doi.org/10.1016/j.wasec.2024.100167

9. Harrington, G. W.; et al. **1992**. Developing a Computer Model to Simulate DBP Formation during Water Treatment. *J. Am. Water Works Assoc*. 84 (11), 78–87. https://doi.org/10.1002/j.1551-8833.1992.tb05886.x

10. Hu, G.; Mian, H. R.; et al. **2023**. Appraisal of Machine Learning Techniques for Predicting Emerging Disinfection Byproducts in Small Water Distribution Networks. *J. Hazard. Mater*. 446, 130633. https://doi.org/10.1016/j.jhazmat.2022.130633

11. Korn, C.; Andrews, R. C.; Escobar, M. D. **2002**. Development of Chlorine Dioxide-Related By-Product Models for Drinking Water Treatment. *Water Res.* 36 (1), 330–342. https://doi.org/10.1016/S0043-1354(01)00194-4

12. Li, Y.; Chen, J.; et al. **2024**. Degradation of the Novel Herbicide Tiafenacil in Aqueous Solution: Kinetics, Various Influencing Factors, Hydrolysis Products Identification, and Toxicity Assessment. *Sci. Total Environ.* 952, 175937. https://doi.org/10.1016/j.scitotenv.2024.175937

13. Ma, Y.; Zhang, X.; et al. **2025**. Simultaneous Control of Chlorite and Micropollutants in Drinking Water through Far-UVC Photolysis. *Environ. Sci. Technol.* 59 (48), 26229–26237. https://doi.org/10.1021/acs.est.5c08408

14. McKelvey, S.; Cheng, Z.; et al. **2025**. Assessment of Regulated and Unregulated Disinfection Byproduct Formation Using Chlorine Dioxide Pre-Oxidation in Pilot Studies and Data-Driven Modeling. *J. Hazard. Mater*. 497, 139729. https://doi.org/10.1016/j.jhazmat.2025.139729

15. Moeini, M.; Abokifa, A. A. **2025**. Physics-Informed Ensemble Machine Learning for Dynamic Water Quality Prediction in Water Distribution Systems. *J. Water Process Eng.* 79, 108841. https://doi.org/10.1016/j.jwpe.2025.108841

16. Mokiienko, A. V. **2025a.** *Chlorine Dioxide. Vol. 1: Chemistry*; National University of Ostroh Academy: Ostroh, Ukraine. (in Ukrainian)

17. Mokiienko, A. V. **2025b.** *Chlorine Dioxide. Vol. 2: Microbiology*; National University of Ostroh Academy: Ostroh, Ukraine. (in Ukrainian)

18. Mokiienko, A. V. **2026**. *Chlorine Dioxide. Vol. 3: Toxicology*; National University of Ostroh Academy: Ostroh, Ukraine. (in Ukrainian)

19. Shumilova, O.; Tockner, K.; et al. **2023**. Impact of the Russia–Ukraine Armed Conflict on Water Resources and Water Infrastructure. Nat. Sustain. 6, 578–586. https://doi.org/10.1038/s41893-023-01068-x

20. Song, B.; Yang, H.; et al. **2023**. Disinfection Byproducts Formed from Oxidation of Pesticide Micropollutants in Water: Precursor Sources, Reaction Kinetics, Formation, Influencing Factors, and Toxicity. *Chem. Eng. J.* 475, 146310. https://doi.org/10.1016/j.cej.2023.146310

21. Sorlini, S.; Biasibetti, M.; et al. **2016**. Modeling and Analysis of Chlorine Dioxide, Chlorite, and Chlorate Propagation in a Drinking Water Distribution System. *J. Water Supply Res. Technol.–AQUA* 65 (8), 597–611.

22. Sorlini, S.; Collivignarelli, C. **2005**. Chlorite Removal with Granular Activated Carbon. *Desalination* 176 (1–3), 255–265. https://doi.org/10.1016/j.desal.2004.10.030

23. Sorlini, S.; Gialdini, F.; et al. **2014**. Influence of Drinking Water Treatments on Chlorine Dioxide Consumption and Chlorite/Chlorate Formation. *Water Res.* 54, 44–52. https://doi.org/10.1016/j.watres.2014.01.038

24. Strokal, V.; Kurovska, A.; Strokal, M. **2023**. More River Pollution from Untreated Urban Waste Due to the Russian–Ukrainian War: A Perspective View. *J. Integr. Environ. Sci.* 20 (1). https://doi.org/10.1080/1943815X.2023.2281920

25. Suquet, J.; Godo-Pla, L.; et al. **2021**. Assessing the Effect of Catchment Characteristics to Enhanced Coagulation in Drinking Water Treatment: RSM Models and Sensitivity Analysis. *Sci. Total Environ.* 799, 149398. https://doi.org/10.1016/j.scitotenv.2021.149398

26. Zalenska, Y. **2023**. Surface Water Quality Assessment for Irrigation Using Harrington's Desirability Function. *Probl. Khim. Stal. Rozv.* 8, 32–39. https://doi.org/10.32782/pcsd-2023-2-5 (in Ukrainian)

27. Zanetti, M.; Godio, A.; et al. **2008**. Chlorine Dioxide By-Products Predictive Models for Drinking Water Oxidation Treatment. *Water Sci. Technol.: Water Supply* 8 (3), 331–338. URL: https://www.academia.edu/108208653/Chlorine_dioxide_by_products_predictive_models_for_drinking_water_oxidation_treatment

28. Zhang, R.; Lin, Z.; et al. **2025**. An In-Situ Strategy of Modified Biochar–Activated Persulfate for Low-Temperature Livestock Wastewater Treatment: Dynamic Reaction Kinetics and Computational Fluid Dynamics-Based Multi-Unit Structure Reactor Design. *Chem. Eng. Process.* 216, 110412. https://doi.org/10.1016/j.cep.2025.110412

ПРОГНОЗУВАННЯ УТВОРЕННЯ ПОБІЧНИХ ПРОДУКТІВ ПІД ЧАС ЗНЕЗАРАЖЕННЯ ВОДИ ДІОКСИДОМ ХЛОРУ: ВІД ЕКСПЕРИМЕНТАЛЬНО-СТАТИСТИЧНИХ ДО ГІБРИДНИХ МОДЕЛЕЙ

Олександр Бондарчук[1], *Аркадій Шахновський*[1], *Лариса Спасьонова*[1], *Андрій Мокієнко*[2]

[1] Київський політехнічний інститут імені Ігоря Сікорського, Bondarchukau@gmail.com

[2] Національний університет «Острозька академія», Mokienkoav56@gmail.com

Забезпечення безпечною питною водою залишається одним із найкритичніших гуманітарних викликів, що постають перед Україною в умовах триваючого збройного конфлікту. Руйнування водної інфраструктури в поєднанні з погіршенням якості води в джерелах зумовили нагальну потребу в адаптивних, науково обґрунтованих інструментах керування технологічним процесом дезінфекції. Діоксид хлору має суттєві переваги над традиційним хлоруванням, включаючи ефективність у широкому діапазоні pH, здатність руйнувати біоплівки та відсутність канцерогенних тригалогенметанів. Однак дезінфекція діоксидом хлору супроводжується утворенням специфічних побічних продуктів окиснення (переважно хлоритів і хлоратів), сумарна концентрація яких суворо обмежена національними та європейськими нормами. Це створює фундаментальний технологічний парадокс між гарантованою інактивацією патогенів та мінімізацією побічних продуктів. У представленому дослідженні запропоновано комп'ютерно-інтегрований модельний підхід до прогнозування та контролю побічних продуктів дезінфекції діоксидом хлору; запропоновано гібридну архітектуру комп'ютерної моделі як основу для адаптивної, інтегрованої в SCADA системи дозування, здатної коригувати подачу діоксиду хлору в режимі реального часу на основі даних датчиків якості води. Для побудови моделей поверхні відгуку другого порядку було реалізовано п'ятифакторний центральний композиційний ротабельний план (pH, температура, доза діоксиду хлору, час контакту, загальний органічний вуглець) для двох цільових показників: частки спожитого діоксиду хлору як міри ефективності дезінфекції та концентрації хлоритів в очищеній воді. Обидві моделі продемонстрували високу статистичну адекватність. Аналіз рівнянь регресії виявив, що pH є домінуючим фактором, який контролює утворення хлоритів, причому вільний член моделі утворення хлоритів наближається до нормативної межі за середніх робочих умов. Це підтверджує, що пасивне підтримання параметрів не може гарантувати дотримання нормативних вимог. Багатокритеріальна оптимізація з використанням функції бажаності Харрінгтона дозволила запропонувати оптимальний робочий режим: pH 6,0; температура 25 °C; доза діоксиду хлору 6,0 мг/дм^3; час контакту 18,2 години; вміст загального органічного вуглецю (ЗОВ) ≤ 10 мг/дм^3. Аналіз Парето-фронту структурував повний набір оптимальних компромісних рішень за трьома операційними режимами (максимальне знезараження, мінімальне утворення побічних продуктів, та режим технологічного компромісу), що надає водоканалам гнучкий інструмент прийняття рішень, адаптований до поточних епідеміологічних пріоритетів. Стохастичне моделювання за методом Монте-Карло дозволило кількісно оцінити сезонну диференціацію ризиків, продемонструвавши, що зниження pH є ефективнішим важелем контролю, ніж зниження дози, для забезпечення відповідності нормам за вмістом хлоритів в умовах літнього складу вихідної води.

Ключові слова: *очищення питної води, дезінфекція діоксидом хлору, утворення хлоритів, побічні продукти дезінфекції, адаптивне дозування, методологія поверхні відгуку, гібридні моделі*