

METHOD FOR DETERMINING THE RELIABILITY OF THE PROBABILISTIC-STATISTICAL MODEL FOR CALCULATING THE OZONATION PROCESS

V. Myronchuk, Yu. Zmievsii, V. Zakharov, L. Kornienko
National University of Food Technologies

Key words:

Ozone
Ozone-gas composition
Precision
Recal
F-score

Article history:

Received 15.07.2020
Received in revised form
30.07.2020
Accepted 14.08.2020

Corresponding author:

V. Zakharov

E-mail:

saharoff.911@gmail.com

ABSTRACT

The paper presents a method for determining the reliability of the probability-statistical model for calculating the ozonation process.

Due to the statistical analysis of frequency ranges and probability distributions of different parameters by classes, it was found that the temperature has a greater effect on the ozonation process than the concentration of ozone in the ozone-gas mixture. The temperature value makes it possible to determine the range of ozone solubility, that is to calculate the efficiency of the ozonation process.

The result of further work was a method for determining the reliability of the probability-statistical model for calculating the ozonation process. Metrics such as Precision and Recall were used to analyze the data processing algorithms. Subsequently, they were used to calculate another metric — F-score. This metric takes into account both completeness and accuracy and allowed to determine the efficiency of the calculation algorithm. The F-score was calculated as the average harmonic between accuracy and completeness and was taken in the range from 0 to 1. The closer the F-score was to 1, the more accurate the calculation results were. Using the proposed method, it was found that the average F-measure for all classes is 0.9069.

The accuracy of calculation of the developed model was established, for the range of ozone solubility within 60...80% accuracy reaches 100%, for other ranges accuracy fluctuates from 83 to 99(%), at average accuracy of 92.25% that in general is a high indicator.

The proposed model allows to determine the efficiency of the ozonation process depending on the set values of temperature and ozone concentration in the ozone-gas mixture.

In the future, the method proposed by the authors to determine the reliability of the probabilistic-statistical model of calculation of the ozonation process can provide constant control over the correctness of the calculations when replenishing the sample data of the model.

МЕТОДИКА ВИЗНАЧЕННЯ ДОСТОВІРНОСТІ ЙМОВІРНІСНО-СТАТИСТИЧНОЇ МОДЕЛІ РОЗРАХУНКУ ПРОЦЕСУ ОЗОНУВАННЯ

В. Г. Мирончук, Ю. Г. Змієвський, В. В. Захаров, Л. В. Корнієнко
Національний університет харчових технологій

У статті представлено методика визначення достовірності ймовірно-статистичної моделі для розрахунку процесу озонування.

Завдяки проведенню статистичного аналізу полігонів частот і розподілів ймовірностей різних параметрів по класах виявлено, що на процес озонування температура впливає більше, ніж концентрація озону в озono-газовій суміші. Значення температури дає змогу визначити діапазон розчинності озону, тобто розраховувати ефективність процесу озонування.

Результатом подальших досліджень стала методика визначення достовірності ймовірно-статистичної моделі для розрахунку процесу озонування. Для аналізу алгоритмів обробки даних використовувались такі метрики, як точність (Precision) і повнота (Recall). В подальшому їх використали для розрахунку ще однієї метрики — F-міри (F-score). Ця метрика враховує і повноту, і точність, тому стало можливим визначити ефективність алгоритму розрахунку. F-міру розраховували як середнє гармонічне між точністю та повнотою та приймали в межах значення від 0 до 1. Чим ближче F-міри знаходились до 1, тим точнішими були результати розрахунку. За допомогою запропонованої методики встановлено, що усереднений за всіма класами показник F-міри дорівнює 0,9069.

Встановлено точність розрахунку розробленої моделі. Так, для діапазону розчинності озону в межах 60...80% точність сягає 100%, для інших діапазонів точність коливається від 83 до 99(%), при середній точності 92,25%, що загалом є високим показником.

Запропонована модель дає змогу визначити ефективність процесу озонування залежно від заданих значень температури та концентрації озону в озono-газовій суміші.

У подальшому запропонована авторами методика для визначення достовірності ймовірно-статистичної моделі розрахунку процесу озонування може забезпечити постійний контроль за правильністю розрахунків при поповненні вибірки даних моделі.

Ключові слова: озон, озono-газова суміш, точність, повнота, F-міра.

Постановка проблеми. Застосовуючи процес озонування, можна видаляти небажані органічні домішки з оброблюваних розчинів і забезпечувати мікробіологічну чистоту технологічного обладнання на підприємствах харчової промисловості [1—4]. Перевагою озонування перед схожими способами обробки (з

допомогою окислення) є його екологічність і безпечність для харчових виробництв [1; 4; 5]. Вже існує цілий ряд підприємств харчової галузі, які застосовують озонування для унеможливлення та знищення плісняви на тарі, упаковці й обладнанні. Також використовують обробку озоном слабоалкогольних напоїв, соків та молока з метою підвищення їхнього терміну придатності [1; 3—7]. На сьогодні не існує чітко прописаного та загальновизнаного способу розрахунку або прогнозування ефективності процесу озонування, тому дослідження в цьому напрямі є актуальними.

У статті [8] представлено створення ймовірно-статистичної моделі розрахунку процесу озонування та запропоновано використовувати алгоритм Баєса для її вирішення. Отримана модель дала змогу проводити розрахунок і підбір необхідних параметрів для роботи станцій озонування, зокрема значення розчинності озону залежно від температури та концентрацію озону в озono-газовій суміші. Однією з головних переваг запропонованої моделі розрахунку є постійне поповнення бази даних, що робить можливим покращення роботи моделі. Водночас додавання нової інформації потребує постійного якісного аналізу розподілення даних по класах для запобігання накладання їх один на одного. Також важливим загалом та особливо для ймовірно-статистичних моделей є оцінка достовірності (адекватності) створеної моделі, оскільки це дає змогу впевнитись у правильності математичних розрахунків.

Мета дослідження: розроблення доступної та універсальної методики визначення достовірності ймовірно-статистичної моделі для розрахунку розчинності озону.

Результати і обговорення. Побудова полігонів відносних частот та їх аналіз. Згідно з [8] розподіл значень розчинності озону було розбито на чотири класи (табл. 1).

Таблиця 1. Поділ діапазону значень розчинності озону на класи

Клас	C ₁	C ₂	C ₃	C ₄
Діапазон значень, %	$60 \leq Y < 80$	$40 \leq Y < 60$	$20 \leq Y < 40$	$0 \leq Y < 20$

Вихідні дані представлено у вигляді бази даних (табл. 2), яка може постійно поповнюватися новою інформацією для розрахунку.

Таблиця 2. Формування класів і параметрів вибірки

№	Температура озono-газової суміші t , °C	Концентрація озону в озono-газовій суміші x , г/м ³	Розчинність озону Y , %	Клас
	5	120	47	C ₂
	0	70	66	C ₁
	10	190	40	C ₂
	30	20	17	C ₄
	25	125	15	C ₄
	0	50	67	C ₂
...

Для якісного аналізу вихідних даних були побудовані полігони відносних частот (рис. 1 та 2) по кожному класу (для різних діапазонів розчинності озону) [2; 3].

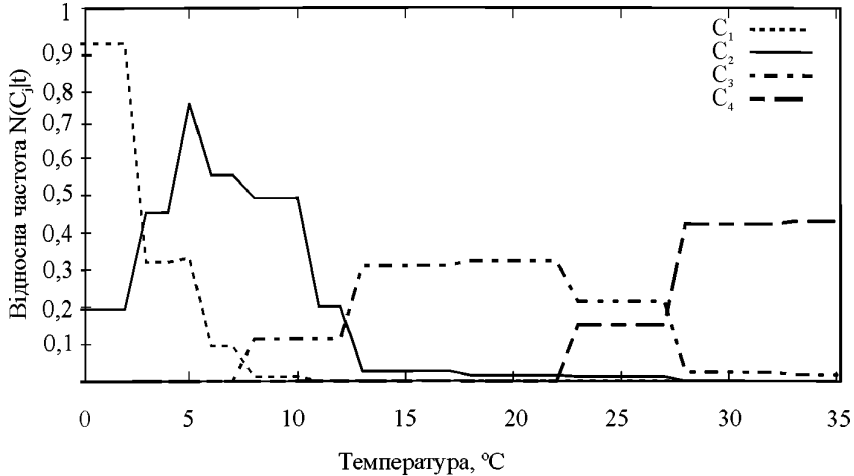


Рис. 1. Полігон відносних частот для температури озono-газової суміші

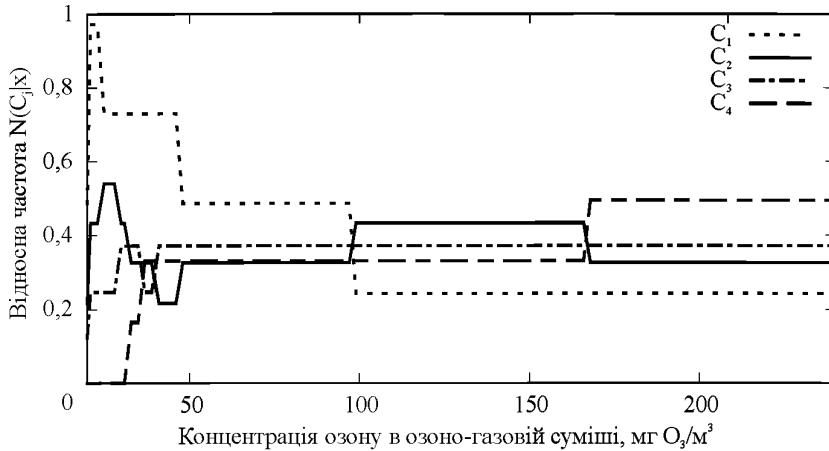


Рис. 2. Полігон відносних частот для концентрації озону в озono-газовій суміші

Зазвичай, для аналізу використовуються згладжені (неперервні) полігони відносних частот [9; 10]. Щоб перейти від дискретного розподілу до неперервного (рис. 3 та 4) використовували згладжування методом Безьє (з параметром $n = 3$, що відповідає кубічним поліномам) за такими співвідношеннями:

$$P(t) = \sum_{i=0}^n P_i \cdot J_{ni}(a), \quad 0 < a < 1, \quad (1)$$

де $J_{ni}(a) = C_n^i \cdot a^i \cdot (1-a)^{n-i}$ — поліноми Бернштейна; a — змінна величина ($a = t$ для розподілів температур та $a = x$ для розподілів концентрацій).

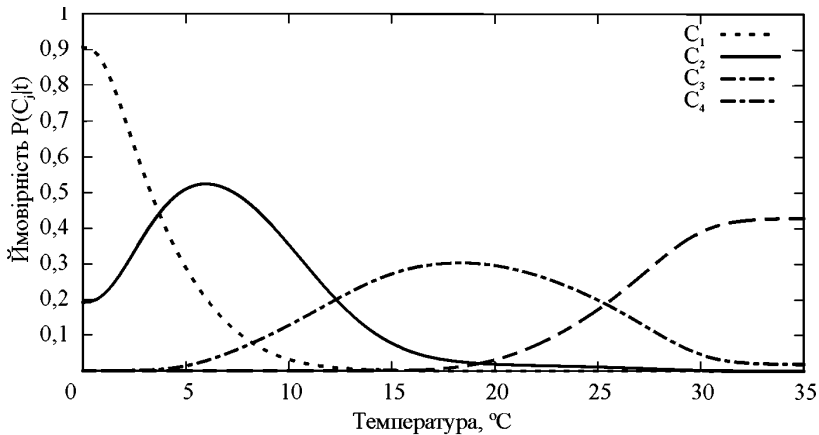


Рис. 3. Згладжений полігон відносних частот для температури озono-газової суміші

Отримавши полігони відносних частот та їх згладжені побудови можна робити певні висновки про якість використаних даних.

Аналіз частотних полігонів і розподілів імовірностей показав (рис. 1 та 3), що кожний з класів має своє значення екстремуму (табл. 3). Можна зробити висновок, що температура має значний вплив на ефективність процесу озонування, а також дає змогу визначити діапазон розчинності озону, тобто розрахувати ефективність процесу озонування.

Другий параметр, концентрація озону в газовій суміші (x), показав нижчу цінність для визначення ефективності процесу озонування (визначення класу). Починаючи з концентрації $x = (110 \sim 240)$ $\text{мг}(\text{O}_2)/\text{м}^3$, імовірність належності процесу до класу C_1 становить $\sim 27\%$ і не змінюється. Розподіл для класу C_1 (рис. 2 та 4) можна пояснити так: при малій концентрації більша кількість озону розчиняється, а при збільшенні концентрації спостерігається поступове зменшення кількості розчиненого озону. Розподіли по класах C_2, C_3 і C_4 для x значно накладаються один на одного і лише в діапазоні $(0 \leq x \leq 60)$ $\text{мг}(\text{O}_2)/\text{м}^3$ чітко виділяється клас C_1 . Незважаючи на це, початкова концентрація озону x може бути застосована як додатковий параметр для розрахунку.

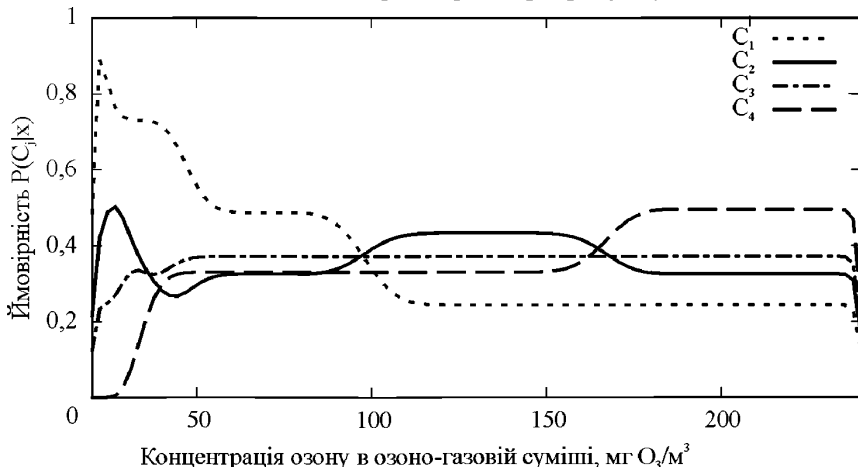


Рис. 4. Згладжений полігон відносних частот для температури озono-газової суміші

Таблиця 3. Вплив значень температури та концентрації озону в озono-газовій суміші на належність до класу

Належність до класу	Діапазон значення температури, °C	Діапазон концентрацій, мг(O ₂)/м ³
C ₁	0...2,5	20... 60
C ₂	3... 10	110... 160
C ₃	15... 22	50... 240
C ₄	30... 35	170... 240

Оцінка достовірності (адекватності) створеної ймовірно-статистичної моделі є важливим заключним етапом при розробці моделі. Одна з найбільш широкоживаних методик визначення точності (адекватності) математичних ймовірно-статистичних моделей є F-міра або в міжнародному позначенні F-score. Цей показник розраховується для кожної множини подій (для кожного класу) за результатами тестування. «Позитивна» оцінка означає, що подія (набір вхідних параметрів або даних) належить до класу C_i, «Негативна» — не належить до класу C_i. У подальшому оцінка моделі порівнюється з певним еталоном, зазвичай з вибіркою експериментальних даних. Вірними прогнозами моделі є ті події, для яких оцінка моделі збігається з експертною оцінкою.

Для аналізу алгоритмів обробки даних використовувались такі метрики, як точність (Precision) та повнота (Recall). Точність у межах класу — це частка подій, що дійсно належать до певного класу відносно всіх подій, які модель віднесла до цього ж класу. Повнота — це частина визначених класифікатором подій, які належать до певного класу відносно всіх подій (даних з вибірки) [9; 10].

Чим більше повнота й точність наближаються до 100%, тим точніше відбувається розрахунок (класифікація), але на практиці максимальна повнота та точність одночасно недосяжні, тому необхідно знаходити певний баланс. F-міра (F-score) — метрика, що враховує і повноту, і точність та дає змогу визначити адекватність (ефективність) алгоритму розрахунку. F-міра розраховується як середнє гармонічне між точністю та повнотою (формула 3) та набуває значення від 0 до 1. При низькій точності або повноті F-міра прямує до 0, при високих показниках — до 1 [9; 10].

$$F_{score} = 2 \frac{Precision \cdot Recall}{Precision + Recall} \quad (3)$$

У свою чергу Precision та Recall розраховуються таким чином:

$$Precision = \frac{TP}{TP + FP} \quad ; \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

де TP — позитивно-позитивна оцінка; TN — позитивно-негативна оцінка; FP — негативно-позитивна оцінка; FN — негативно-негативна оцінка.

Було проведено тестування математичної моделі для наборів (комбінацій) вхідних параметрів: концентрація озону x від 20 до 240 (г/м³), з кроком 1 г/м³

та для температур t від 0 до 35(°C), з кроком 1°C. Результати тестування по кожному класу приведені у вигляді матриць помилок у табл. 4.

Таблиця 4. Матриця помилок при тестуванні по кожному класу

		Оцінка за експериментальними даними	
Клас С ₁			
Розрахунок моделі (оцінка моделі)		Позитивна	Негативна
	Позитивна	298	0
	Негативна	31	1871
Клас С ₂			
Розрахунок моделі (оцінка моделі)		Позитивна	Негативна
	Позитивна	698	104
	Негативна	41	1357
Клас С ₃			
Розрахунок моделі (оцінка моделі)		Позитивна	Негативна
	Позитивна	570	114
	Негативна	77	1439
Клас С ₄			
Розрахунок моделі (оцінка моделі)		Позитивна	Негативна
	Позитивна	412	4
	Негативна	73	1711

Використовуючи матрицю помилок та формули (3)—(5) було розраховано значення Precision, Recall та F-міру для кожного із класів.

Таблиця 5. Сумарні результати аналізу ймовірісно-статистичної моделі з використанням F-міри

Клас	Precision (точність), %	Recall (повнота), %	F-міра
C ₁	100,00	91,00	0,9506
C ₂	87,00	94,00	0,9059
C ₃	83,00	88,00	0,8565
C ₄	99,00	85,00	0,9145
Середнє значення по всіх класах	92,25	89,50	0,9069

За аналізом статистичних метрик для представленої ймовірісно-статистичної моделі (табл. 5) при зміні вхідних параметрів від 20 до 240 (г/м³) для концентрації озону в озono-газовій суміші x з кроком 1 г/м³ і температури t від 0 до 35(°C) з кроком 1°C можна зробити такі висновки: для діапазону розчинності озону в межах від 60 до 80% (клас С₁) точність класифікації сягає 100%. Для всіх інших класів точність коливається від 83 до 99(%) при середній точності 92,25%, що є високим показником. Узагальнений та усереднений за всіма класами показник F-міра дорівнює 0,9069.

Висновки

Застосування ймовірісно-статистичної моделі розрахунку процесу озонування з використанням полігонів частот підтвердило доцільність вибору температури й концентрації озону в озono-газовій суміші як параметрів для розрахунку значення розчинності озону. Достовірність і адекватність запропонованої моделі

з використанням F-міри показала високу точність розрахунку розчинності озону, при цьому усереднений показник F-міра дорівнює 0,9069.

Література

1. Degremont. Технический справочник по обработке воды. Том 1 и Том 2. / Andriamirado L. и др. Спб., 2007. 1798 с.
2. Clark P. J. Ozone Cure for Some Sanitation Problems. *Food Technology*. 2004. № 58. P. 75—76.
3. Koros W. J., Ma Y. H., Shimidzu T. Terminology for membranes and membrane processes (IUPAC Recommendations). *Pure and Applied Chemistry*. 1996. Vol. 68, is. 7. P. 1479—1489.
4. Worldwide Ozone Capacity for Treatment of Drinking Water and Wastewater: A Review / B. L. Loeb and all. *Ozone: Science & Engineering: The Journal of the International Ozone Association*, 2012. № 34. P. 64—77.
5. Shigezo N., Takahara H. Ozone Contribution in Food Industry in Japan. *Ozone: Science and Engineering*. 2006. № 28. P. 425—429.
6. Efficacy and Fumigation Characteristics of Ozone in Stored Maize / R. A. Pandiselvam and all. *Journal of Stored Products Research*. 2001. № 37, P. 371—382.
7. Khadre N. A., Yousef A. E., Kim J. G. Microbiological Aspects of Ozone Applications in Food: A Review. *Journal of Food Science*. 2001. № 66. P. 1242—1252.
8. Захаров В. В., Устінов О. А., Змієвський Ю. Г., Мирончук В. Г. Застосування алгоритму наївного басового класифікатора для розрахунку та прогнозування процесів озонування. *Наукові праці Національного університету харчових технологій*. 2018. Том 24, № 5, Ч. 1. С. 91—98.
9. Langseth H., Nielsen T. D. Classification using Hierarchical Naive Bayes models. *Mach Learn*. 2006. № 63. С. 135—159.
10. Ковтун Н. В. Теорія статистики. К.: Знання, 2012. 399 с.