



УДК 531.732:006.91

# Вплив порогових значень на стандартну невизначеність типу А при вимірюваннях масової витрати рідини

В.О. Ащепков<sup>1,2</sup>, Д.Ю. Бяллович<sup>1,2</sup>, В.В. Склярів<sup>1,2</sup>

<sup>1</sup> Харківський національний університет радіоелектроніки, просп. Науки, 14, 61166, Харків, Україна

<sup>2</sup> Національний науковий центр "Інститут метрології", вул. Мירוносицька, 42, 61002, Харків, Україна  
ashhepkovvalera@gmail.com

## Анотація

У метрологічних дослідженнях одним із ключових завдань є точне виявлення та усунення викидів у вимірювальних даних, оскільки їх наявність негативно вплине на точність результатів. Раніше проведені нами дослідження були спрямовані на виявлення та усунення викидів за допомогою різних статистичних методів, таких як метод міжквартильного розмаху (IQR) і метод медіанного абсолютного відхилення (MAD), а також методу машинного навчання ізольованого лісу (Isolation Forest). Ці методи показали свою ефективність за певних умов, дозволяючи значно знизити невизначеність типу А та покращити точність вимірювань.

Однак, попри досягнутий прогрес, залишається невирішеним питання визначення оптимальних порогових значень, при яких невизначеність вимірювань стає мінімальною. Це особливо важливо в контексті завдань, де навіть незначне збільшення невизначеності може суттєво вплинути на результати.

Основна увага приділяється вивченню взаємозв'язку між пороговими значеннями та змінами в невизначеності, а також аналізу можливих причин цих змін. Дослідження спрямоване на виявлення оптимальних умов обробки даних, за яких досягається мінімальна невизначеність вимірювань, що є важливим для підвищення точності метрологічних досліджень.

**Ключові слова:** метрологічні дослідження; вимірювання; викиди; аномалії; невизначеність; похибка; порогові значення; метод ізольованого лісу; робастні методи.

Отримано: 28.08.2024

Відредаговано: 23.09.2024

Схвалено до друку: 26.09.2024

## 1. Вступ

У цьому дослідженні використовувалися дані, отримані в результаті метрологічних вимірювань на державному первинному еталоні одиниці об'ємної та масової витрати рідини, об'єму та маси рідини, що протікає по трубопроводу (ДЕТУ 03-04-04). Метою цих вимірювань було визначення метрологічних характеристик коріюлісових витратомірів при різних рівнях масової витрати рідини. Отримані дані включали вибірки значень, що містили як нормальні вимірювання, так і можливі викиди, здатні суттєво вплинути на точність результатів.

Для обробки даних використовувалися кілька статистичних методів, таких як метод міжквартильного розмаху (IQR), метод медіанного абсолютного відхилення (MAD) та алгоритм ізольованого лісу (Isolation Forest). Ці методи були

застосовані для виявлення та усунення викидів з метою зниження стандартної невизначеності типу А [1–2]. Проте під час аналізу було виявлено, що зміна порогових значень цих методів впливає на значення невизначеності. Тому основною метою цього дослідження є вивчення впливу порогових значень на стандартну невизначеність типу А в метрологічних даних.

## 2. Результати розрахунків статистичними методами

Для кожного коріюлісового витратоміра були проведені вимірювання при трьох різних рівнях витрати рідини, що дозволило отримати три набори даних для кожного з трьох витратомірів. Розмір вибірок варіювався від 16 до 33 значень, що забезпечило достатню репрезентативність даних для проведення подальших статистичних аналізів.

Для порівняння стандартної невизначеності вимірювань за типом А, до та після виключення викидів у межах однієї вибірки, використовувалась така формула [3]:

$$u_A(X_j) = \frac{s}{\sqrt{n_0}}, \quad (1)$$

де  $n_0$  – об’єм вибірки до виключення викидів;  
 $s$  – середнє квадратичне відхилення вибірки, яке розраховується за формулою [4]:

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)}}, \quad (2)$$

де  $X_i$  – значення вимірювання;  
 $\bar{X}$  – середнє значення вибірки;  
 $n$  – кількість вимірювань.

Загальноприйнятими порогами для робастних методів є такі:

- метод міжквартильного розмаху: 1,5;
- метод медіанного абсолютного відхилення: 3.

З кроком 0,1 та 0,2 від загальноприйнятих до 0 було проведено розрахунки невизначеності після виключення викидів двома методами.

Розрахунки наведено в табл. 1–2.

Таблиця 1

Результати розрахунку стандартної невизначеності вимірювань типу А при різних порогових значеннях методом міжквартильного розмаху

Порогове значення	Точка масової витрати								
	Витратомір №1			Витратомір №2			Витратомір №3		
	45 т/г	25 т/г	5 т/г	5 т/г	2,5 т/г	1 т/г	1 т/г	0,5 т/г	0,1 т/г
1,4	0,0308	0,0300	0,0174	0,0158	0,0251	0,0070	0,0060	0,0167	0,0476
1,3	0,0308	0,0300	0,0174	0,0158	0,0251	0,0060	0,0060	0,0167	0,0476
1,2	0,0308	0,0300	0,0174	0,0158	0,0251	0,0054	0,0060	0,0167	0,0476
1,1	0,0308	0,0300	0,0174	0,0158	0,0251	0,0047	0,0060	0,0167	0,0476
1	0,0308	0,0300	0,0145	0,0158	0,0251	0,0047	0,0060	0,0167	0,0414
0,9	0,0308	0,0300	0,0145	0,0158	0,0251	0,0047	0,0060	0,0167	0,0349
0,8	0,0308	0,0300	0,0145	0,0158	0,0251	0,0047	0,0048	0,0155	0,0349
0,7	0,0308	0,0300	0,0145	0,0140	0,0251	0,0047	0,0048	0,0155	0,0349
0,6	0,0277	0,0264	0,0117	0,0126	0,0251	0,0043	0,0048	0,0155	0,0349
0,5	0,0239	0,0264	0,0117	0,0126	0,0213	0,0041	0,0048	0,0148	0,0306
0,4	0,0239	0,0264	0,0117	0,0126	0,0202	0,0041	0,0048	0,0144	0,0306
0,3	0,0239	0,0221	0,0107	0,0116	0,0163	0,0032	0,0048	0,0139	0,0274
0,2	0,0308	0,0300	0,0174	0,0158	0,0251	0,0070	0,0060	0,0167	0,0476
0,1	0,0308	0,0300	0,0174	0,0158	0,0251	0,0060	0,0060	0,0167	0,0476

Таблиця 2

Результати розрахунку стандартної невизначеності вимірювань типу А при різних порогових значеннях методом медіанного абсолютного відхилення

Порогове значення	Точка масової витрати								
	Витратомір №1			Витратомір №2			Витратомір №3		
	45 т/г	25 т/г	5 т/г	5 т/г	2,5 т/г	1 т/г	1 т/г	0,5 т/г	0,1 т/г
2,8	0,0207	0,0300	0,0145	0,0158	0,0251	0,0047	0,0060	0,0167	0,0476
2,6	0,0207	0,0300	0,0145	0,0158	0,0251	0,0047	0,0060	0,0155	0,0419
2,4	0,0207	0,0300	0,0145	0,0158	0,0251	0,0047	0,0048	0,0155	0,0349
2,2	0,0136	0,0264	0,0117	0,0158	0,0213	0,0043	0,0048	0,0155	0,0349
2	0,0136	0,0264	0,0117	0,0126	0,0186	0,0043	0,0048	0,0148	0,0349
1,8	0,0136	0,0221	0,0117	0,0126	0,0186	0,0041	0,0048	0,0144	0,0349
1,6	0,0136	0,0221	0,0117	0,0126	0,0186	0,0032	0,0048	0,0139	0,0274
1,4	0,0136	0,0221	0,0107	0,0126	0,0153	0,0032	0,0048	0,0120	0,0274
1,2	0,0132	0,0214	0,0107	0,0111	0,0116	0,0031	0,0045	0,0112	0,0247
1	0,0109	0,0181	0,0086	0,0093	0,0105	0,0028	0,0035	0,0097	0,0167
0,8	0,0074	0,0123	0,0060	0,0071	0,0086	0,0021	0,0026	0,0076	0,0138
0,6	0,0065	0,0113	0,0041	0,0048	0,0086	0,0015	0,0013	0,0052	0,0138
0,4	0,0035	0,0007	0,0041	0,0038	0,0036	0,0009	0,0005	0,0034	0,0068
0,2	0,0035	0,0007	0,0002	0,0038	0,0035	0,0004	0,0005	0,0012	0,0061

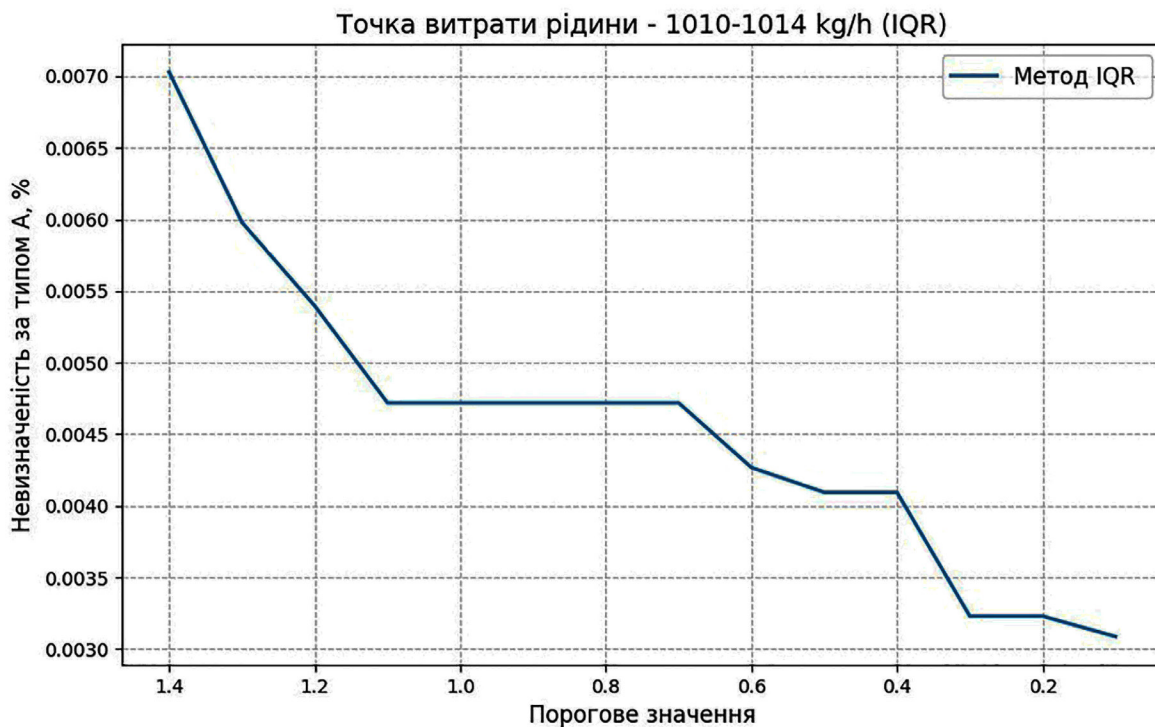


Рис. 1. Знаходження викидів при різних порогових значеннях методом IQR на точці масової витрати 1 т/г

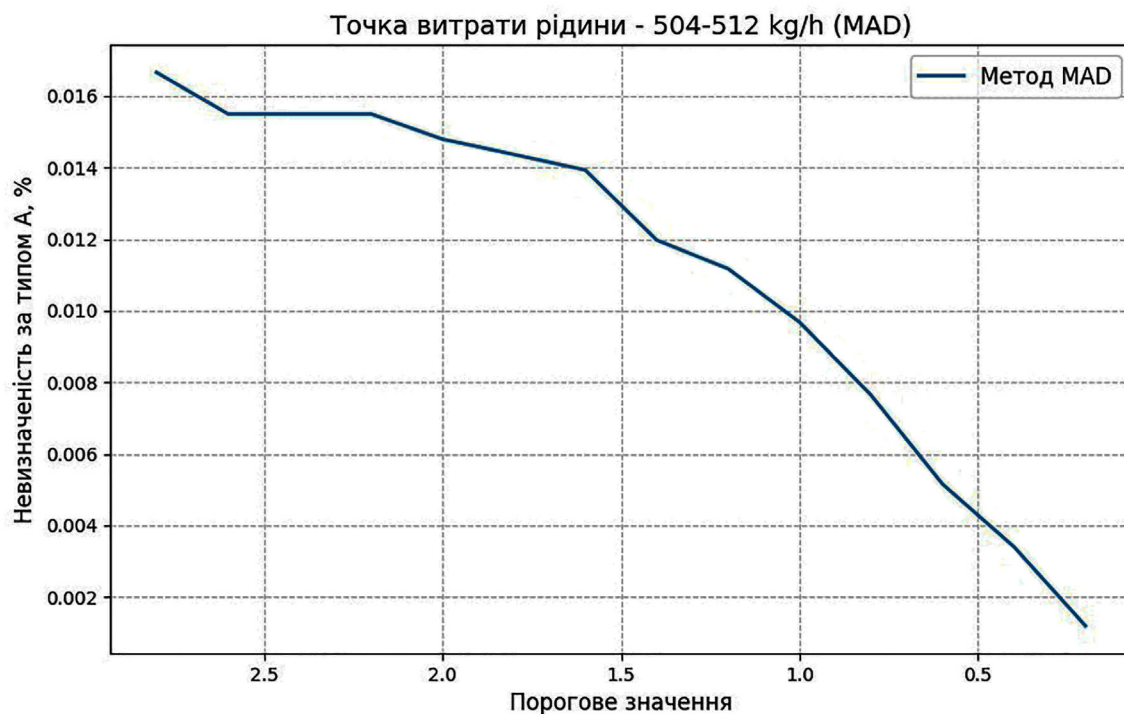


Рис. 2. Знаходження викидів при різних порогових значеннях методом MAD на точці масової витрати 0,5 т/г

Результати проведених розрахунків показали, що зміна порогових значень у всіх вибірках призводила до зменшення стандартної невизначеності (рис. 1–2). Це відбувалося через те, що поступове зниження порогу для виключення викидів зменшувало варіативність даних. Проте, продовжуючи знижувати порогові значення, ми ризикуємо виключити не лише аномальні, але й нормальні значення, що може призвести до

втрати важливої інформації та, як наслідок, до неправильних висновків.

**Метод міжквартильного розмаху (IQR)** визначає інтерквартильний діапазон, встановлюючи межі на рівні першого (25%) і третього (75%) квартилів даних. Цей метод призначений для виявлення даних, які виходять за межі основного масиву значень і можуть вважатися викидами. Оскільки межі визначаються на основі квартилів, метод IQR

Результати розрахунку стандартної невизначеності вимірювань типу А при різних порогових значеннях

Кількість припущених аномалій,%	Точка масової витрати								
	Витратомір №1			Витратомір №2			Витратомір №3		
	45 т/г	25 т/г	5 т/г	5 т/г	2,5 т/г	1 т/г	1 т/г	0,5 т/г	0,1 т/г
0,0	0,0268	0,0446	0,0139	0,0153	0,0232	0,0099	0,0171	0,0153	0,0459
0,05	0,0268	0,0446	0,0139	0,0153	0,0232	0,0089	0,0171	0,0146	0,0459
0,1	0,0231	0,0289	0,0131	0,0140	0,0216	0,0064	0,0117	0,0137	0,0403
0,15	0,0199	0,0253	0,0131	0,0121	0,0201	0,0059	0,0089	0,0135	0,0335
0,2	0,0196	0,0253	0,0102	0,0111	0,0201	0,0055	0,0057	0,0140	0,0335
0,25	0,0196	0,0211	0,0102	0,0111	0,0190	0,0049	0,0046	0,0135	0,0293
0,3	0,0173	0,0181	0,0100	0,0106	0,0182	0,0037	0,0043	0,0128	0,0261
0,35	0,0170	0,0181	0,0104	0,0097	0,0133	0,0035	0,0043	0,0127	0,0239
0,4	0,0109	0,0174	0,0104	0,0101	0,0133	0,0033	0,0039	0,0130	0,0239
0,45	0,0103	0,0139	0,0105	0,0106	0,0109	0,0029	0,0037	0,0127	0,0212
0,5	0,0103	0,0139	0,0107	0,0106	0,0098	0,0028	0,0029	0,0120	0,0190

схильний концентруватися на даних, що зосереджені навколо середніх значень [5].

**Метод медіанного абсолютного відхилення (MAD)**, на відміну від методу IQR, орієнтований на центральне значення вибірки – медіану. Він обчислює абсолютне відхилення кожного значення від медіани та визначає медіанне відхилення як міру розкиду даних. Цей метод також схильний до фокусування на центральній тенденції, оскільки базується на відстані від середини вибірки, що робить його чутливим до концентрації даних навколо медіани [6–7].

Подібно до стандартної невизначеності типу А, яка розраховується на основі середніх значень, обидва ці методи схильні акцентувати увагу на центральних тенденціях вибірки. Таким чином, зменшуючи порогові значення для виявлення та усунення викидів, ми знижуємо вплив значень, які збільшують невизначеність. Однак це може призвести до того, що, знижуючи невизначеність, ми також виключаємо нормальні значення, що негативно позначиться на достовірності результатів [8–10].

### 3. Результати розрахунків методом ізольованого лісу

Алгоритм ізольованого лісу (Isolation Forest) ґрунтується на принципі ізоляції аномальних точок даних за допомогою побудови випадкових дерев. У цьому методі кожне дерево створюється шляхом випадкового поділу даних, і чим легше ізолювати певну точку, тим більша ймовірність того, що вона є викидом. На відміну від традиційних методів, де порогові значення визначаються на основі статистичних параметрів вибірки, в ізольованому лісі порогове значення встановлюється на основі частки аномальних даних, яку користувач вважає допустимою.

Значення порогу зазвичай задається у вигляді відсотка від загальної кількості даних, що дозволяє користувачу контролювати чутливість алгоритму до виявлення викидів. У цьому дослідженні розрахунки стандартної невизначеності типу А було проведено з кроком 5% від 0 до максимального значення 50%. Отримані результати наведено в табл. 3.

За результатами розрахунків було встановлено, що в деяких вибірках спостерігається збільшення стандартної невизначеності типу А при досягненні певного порогового значення (рис. 3). Це може свідчити про те, що починають виключатися нормальні значення, тобто досягнуто скупчення нормальних точок, або про недостатню репрезентативність вибірки.

Проте в інших вибірках спостерігалось планомірне зниження невизначеності навіть при поступовому зниженні порогу.

### 4. Аналіз результатів дослідження

За результатами проведеного дослідження було виявлено пряму залежність між встановленим пороговим значенням і стандартною невизначеністю типу А. Незалежно від застосованого методу, чи то робастні методи, такі як метод міжквартильного розмаху (IQR) або метод медіанного абсолютного відхилення (MAD), було виявлено однаковий принцип роботи – визначення діапазону, в межах якого значення вважаються нормальними. У випадку методу міжквартильного розмаху (IQR) межі діапазону визначаються за краями, тоді як у методі медіанного абсолютного відхилення (MAD) вихідною точкою є медіана, навколо якої будується діапазон.

Крім того, аналогічні тенденції спостерігаються при використанні методів, заснованих на середніх значеннях, таких як метод 3-сигм, а також при



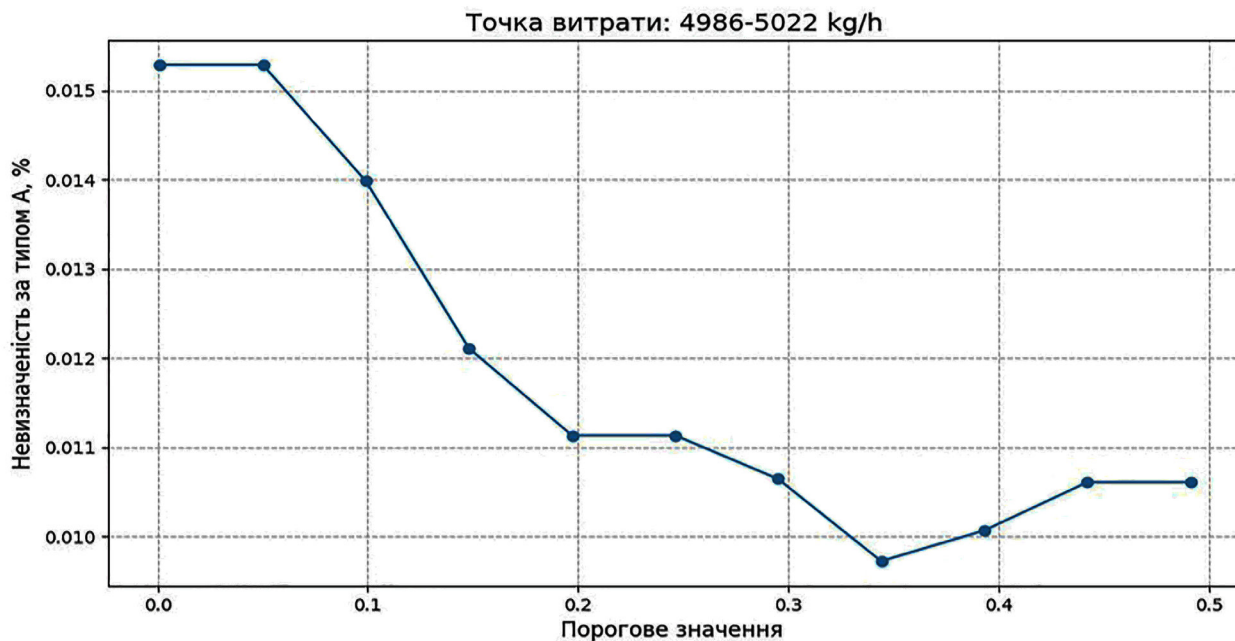


Рис. 3. Знаходження викидів при різних порогових значеннях методом ізольованого лісу на точці масової витрати 5 т/г витратоміра №2

застосуванні статистичних критеріїв і тестів. Усі ці методи, незважаючи на відмінності у підходах, мають спільну концепцію: вони спираються на статистичні параметри та центральні тенденції даних для визначення меж діапазону нормальних значень.

Усі вищезгадані методи призводять до схожих результатів і висновків: при збільшенні чутливості методу невизначеність вимірювань зменшується. Це відбувається через те, що більш чутливі методи виключають більшу кількість даних, потенційно усуваючи викиди, і таким чином скорочують варіативність даних. Однак за відсутності чітко встановленого порогового значення цей процес може нагадувати фільтрацію даних, а не точне визначення викидів і нормальних значень. Таким чином, важливість правильного вибору порогу стає ключовою для досягнення балансу між фільтрацією даних і коректним виявленням викидів, що, у свою чергу, безпосередньо впливає на рівень невизначеності.

**Метод ізольованого лісу** належить до групи методів виявлення викидів, заснованих на відстанях. Ці методи аналізують різницю між точками у багатовимірному просторі й на цій основі визначають, наскільки легко або складно ізолювати ту чи іншу точку даних. Зокрема, ізольований ліс буде набір випадкових дерев, де кожна точка розділяється випадковим чином, і чим легше ізолювати точку, тим вища ймовірність, що вона є викидом.

Метод ізольованого лісу, як показали дослідження, також залежить від порогу, що вказує на необхідність розробки підходів, спрямованих на знаходження достовірних значень, а не просто на встановлення порогу як такого. На відміну

від статистичних методів, які використовують фіксований поріг для виявлення викидів, ізольований ліс має перевагу в тому, що оперує ступенем аномальності. Це дозволяє використовувати більш гнучкі стратегії для виявлення викидів і визначення моменту, коли виключення даних починає зачіпати нормальні значення.

Завдяки використанню ступеня аномальності метод ізольованого лісу надає можливість не просто встановлювати поріг, а виявляти точку, в якій виключення даних починає змінювати загальну тенденцію. Замість налаштування жорсткого порогу можна аналізувати зсув в аномальності, що дозволить точніше визначити момент, коли виключення точок починає негативно впливати на вибірку, зачіпаючи нормальні дані.

Таким чином, замість розробки універсального порогу, який може не підходити для всіх наборів даних, необхідно розробити підхід, спрямований на визначення достовірних значень у кожному конкретному випадку.

## 5. Висновки

У ході проведеного дослідження було проаналізовано вплив різних порогових значень на стандартну невизначеність типу А у метрологічних даних. Для виявлення та усунення викидів використовувалися кілька статистичних методів, зокрема метод міжквартильного розмаху (IQR), метод медіанного абсолютного відхилення (MAD) та алгоритм ізольованого лісу (Isolation Forest).

Статистичні методи, такі як IQR і MAD, показали зміну рівня невизначеності при різних значеннях порогу. Хоча в обох методах спостерігалася залежність між порогом та

невизначеністю, навіть у методі медіани виявлялися зміни, що вказує на труднощі у встановленні універсального порогу, який би підходив для всіх вибірок.

Алгоритм ізольованого лісу також залежить від порогового значення, проте його перевага полягає у використанні ступеня аномальності для аналізу. Це дозволяє не тільки враховувати поріг, а й аналізувати момент, коли виключення даних починає впливати на нормальні значення. Незначне

збільшення невизначеності спостерігалось лише в кількох вибірках, що підкреслює необхідність індивідуального підходу для кожного набору даних.

Таким чином, для забезпечення максимальної точності вимірювань важливо відмовлятися від жорсткого встановлення порогів і розробляти адаптивні підходи, що дозволять зберегти баланс між виключенням викидів і мінімізацією невизначеності для кожної конкретної вибірки.

## The influence of threshold values on Type A standard uncertainty in mass flow rate measurements for liquids

V. Ashchepkov<sup>1,2</sup>, D. Byallovich<sup>1,2</sup>, V. Skliarov<sup>1,2</sup>

<sup>1</sup> Kharkiv National University of Radio Electronics, Nauky Ave., 14, 61166, Kharkiv, Ukraine

<sup>2</sup> National Scientific Centre "Institute of Metrology", Myronosytska Str., 42, 61002, Kharkiv, Ukraine  
ashhepkovvalera@gmail.com

### Abstract

In metrological studies, the minimization of Type A standard measurement uncertainty, which can be significantly increased by the presence of outliers in the data, is an essential task. Our previous research focused on detecting and eliminating outliers using various statistical methods, such as the Interquartile Range (IQR) method, the Median Absolute Deviation (MAD) method, as well as the machine learning-based Isolation Forest method. These methods have proven effective under certain conditions, significantly reducing Type A uncertainty and improving the measurement accuracy.

However, despite the progress made, the issue of determining the optimal threshold values at which measurement uncertainty may well be minimized remains unresolved. This is especially crucial for tasks where even a slight increase in the uncertainty may significantly affect the results.

The main focus is on studying the relationship between threshold values and changes in the measurement uncertainty, as well as analysing possible causes of these changes. The study is aimed at identifying the optimal data processing conditions under which minimal measurement uncertainty can be achieved, which is of paramount importance for improving the accuracy of metrological studies.

**Keywords:** metrological studies; measurements; outliers; anomalies; uncertainty; error; threshold values; Isolation Forest method; robust methods.

### Список літератури

1. Ащепков В.О. Використання моделі Isolation Forest для виявлення аномалій у даних вимірювань. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 1 (27). doi: 10.30837/ITSSI.2024.27.236
2. Ащепков В.О. Дослідження метрологічних характеристик державного первинного еталона одиниці об'ємної та масової витрати рідини при підготовці до участі у міжнародних звіреннях. *Український метрологічний журнал*. 2024. № 1. С. 31–37. doi: 10.24027/2306-7039.1.2024.300937
3. Rostron P.D., Fearn T., Ramsey M.H. Comparing uncertainties – Are they really different? *Accred Qual Assur*, 2022, vol. 27, pp. 133–142. doi: 10.1007/s00769-022-01501-2
4. Zakharov I., Serhienko M., Chunikhina T. Measurement uncertainty evaluation by kurtosis method at calibration of a household water meter. *XXX International Scientific Symposium "Metrology and Metrology Assurance" (MMA)*, 2020, pp. 83–86. doi: 10.1109/MMA49863.2020.9254260
5. Wan X., Wang W., Liu J. et al. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol*, 2014, no. 14 (135). doi: 10.1186/1471-2288-14-135

6. Hozo S.P., Djulbegovic B., Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol*, 2005, no. 5 (13). doi: 10.1186/1471-2288-5-13
7. Blom G. Statistical Estimates and Transformed Beta Variables. New York, John Wiley and Sons, 1958. 176 p.
8. Потаніна Т.В., Михайленко І.В. Дослідження вибірок експериментальних даних на наявність викидів: порівняння методів. *Інтегровані технології та енергозбереження*. 2023. № 3. doi: 10.20998/2078-5364.2023.3.07
9. Wada K. Outliers in official statistics. *Japanese Journal of Statistics and Data Science*, 2020, vol. 3, pp. 669–691. doi:10.1007/s42081-020-00091-y
10. Orellana M., Cedillo P. Outlier Detection with Data Mining Techniques and Statistical Methods. *Proceedings of 2019 International Conference on Information Systems and Computer Science (INCISCOS)*. Quito, Ecuador, 2019, pp. 51–56. doi: 10.1109/INCISCOS49368.2019.00017
- 2024, no. 1, pp. 31–37 (in Ukrainian). doi: 10.24027/2306-7039.1.2024.300937
3. Rostron P.D., Fearn T., Ramsey M.H. Comparing uncertainties – Are they really different? *Accred Qual Assur*, 2022, vol. 27, pp. 133–142. doi: 10.1007/s00769-022-01501-2
4. Zakharov I., Serhiienko M., Chunikhina T. Measurement uncertainty evaluation by kurtosis method at calibration of a household water meter. *XXX International Scientific Symposium “Metrology and Metrology Assurance” (MMA)*, 2020, pp. 83–86. doi: 10.1109/MMA49863.2020.9254260
5. Wan X., Wang W., Liu J. et al. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol*, 2014, no. 14 (135). doi: 10.1186/1471-2288-14-135
6. Hozo S.P., Djulbegovic B., Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol*, 2005, no. 5 (13). doi: 10.1186/1471-2288-5-13
7. Blom G. Statistical Estimates and Transformed Beta Variables. New York, John Wiley and Sons, 1958. 176 p.
8. Potanina T.V., Mykhaylenko I.V. Doslidzhennia vybirok eksperymentalnykh danykh na naiavnist vykydiv: porivniannia metodiv [Examination of Experimental Data Samples for the Presence of Outliers: Comparison of Methods]. *Integrated Technologies and Energy Saving*, 2023, no. 3 (in Ukrainian). doi: 10.20998/2078-5364.2023.3.07
9. Wada K. Outliers in official statistics. *Japanese Journal of Statistics and Data Science*, 2020, vol. 3, pp. 669–691. doi:10.1007/s42081-020-00091-y
10. Orellana M., Cedillo P. Outlier Detection with Data Mining Techniques and Statistical Methods. *Proceedings of 2019 International Conference on Information Systems and Computer Science (INCISCOS)*. Quito, Ecuador, 2019, pp. 51–56. doi: 10.1109/INCISCOS49368.2019.00017

#### References

1. Ashchepkov V.O. Vykorystannia modeli Isolation Forest dlia vyavlennia anomalii u danykh vymiriuvan [The use of the Isolation Forest model for anomaly detection in measurement data]. *Innovative technologies and scientific solutions for industries*, 2024, no. 1 (27) (in Ukrainian). doi: 10.30837/ITSSI.2024.27.236
2. Aschepkov V.O. Doslidzhennia metrolohichnykh kharakterystyk derzhavnoho pervynnoho etalona odyntsi obiemnoi ta masovoi vytraty ridyny pry pidhotovtsi do uchasti u mizhnarodnykh zvirenniakh [Study of metrological characteristics of the state primary measurement standard of volume flow and mass consumption of liquid in preparation for participation in international comparisons]. *Ukrainian Metrological Journal*,