DOI 10.34185/1562–9945–5–166–2026–11
УДК 681.128+681.518.2

Д.В. Іващев, В.В. Герасимов

КОМПЛЕКСНА МЕТОДОЛОГІЯ ОБРОБКИ ТЕЛЕМЕТРИЧНИХ ДАНИХ РІВНЯ ПАЛИВА МАНЕВРОВИХ ЛОКОМОТИВІВ: НЕЛІНІЙНИЙ АНАЛІЗ ШУМУ ТА ДЕТЕКЦІЯ ПОДІЙ ЗАПРАВКИ

Анотація. Запропоновано комплексну методологію обробки телеметричних даних рівня палива маневрових локомотивів, яка об'єднує в один конвеєр сегментацію, адаптивне згладжування з детекцією заправок, класифікацію нелінійного шуму та порівняння методів його усунення. На даних шести локомотивів виявлено 280 подій заправки та класифіковано шум понад 27 тисяч сегментів, де переважає кольоровий шум. Запропоновано формалізований показник довіри до кожної виявленої заправки. Показано, що приріст відношення сигнал/шум, обчислений на ізольованому шумі, не розрізняє якісної фільтрації й підлаштування під шум, тому для впровадження придатні лише фільтр ковзного середнього та фільтр Калмана.
Ключові слова: обробка часових рядів, цифрова фільтрація сигналу, адаптивне диференційоване згладжування, вейвлет–порогова обробка, робастне оцінювання, детекція подій, класифікація шуму, відношення сигнал/шум.

Постановка проблеми. Маневрові локомотиви є ключовим елементом промислових залізничних вузлів, і точний облік витрат і заправок палива дозволяє оцінювати реальну витрату та формувати достовірну звітність. Сучасні системи телеметричного моніторингу здатні фіксувати рівень палива, швидкість руху та потужність дизель–генераторної установки з високою періодичністю, що в масштабах від кількох місяців до двох з половиною років експлуатації дає масиви обсягом у кілька мільйонів вимірювань на кожен об'єкт спостереження.

Разом з тим, зростання обсягу даних не супроводжується автоматичним зростанням їх якості. Показання датчиків рівня палива містять значну шумову складову, а телеметричний потік переривається розривами зв'язку різної тривалості – від коротких, тривалістю кілька хвилин, до тривалих пауз, що в окремих випадках досягають кількох місяців. Усе це разом ускладнює як пряме застосування стандартних статистичних методів до необробленого ряду, так і автоматичне виявлення подій, що мають практичне значення, зокрема заправок пального.

Попередні дослідження авторів розв'язували ці задачі поетапно, кожна робота – на власній вибірці. Спочатку було запропоновано алгоритм сегментації даних на однорідні за режимом роботи ділянки [1]. Далі – алгоритм диференційованого адаптивного згладжування з детекцією заправок [2], що дозволяє очистити сигнал і визначити обсяг залитого палива без спотворення фронту підйому рівня. Окремо запропоновано методику

© Іващев Д.В., Герасимов В.В., 2026
Copyright for this paper by its authors. Use permitted under License CC BY 4.0.
ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

класифікації шуму за спектральним і статистичним типом [3]; її зміст детальніше розглянуто в наступному розділі.

Ускладнює обробку й характер самого шуму: за результатами [3] шум рівня палива переважно є кольоровим, а не білим, тобто його сусідні відліки статистично залежні. Тому прості порогові правила чи згладжування з фіксованим вікном, ефективні для білого шуму, можуть спотворювати корисний сигнал.

Практичне застосування цих трьох методик на новій, ширшій вибірці даних виявило потребу в трьох додаткових кроках. По–перше, окремі методики варто об'єднати в єдиний відтворюваний програмний конвеєр обробки, а не застосовувати послідовно як незалежні інструменти. По–друге, для оцінки надійності детекції заправок на великій вибірці, де ручна верифікація кожної події є практично неможливою, потрібен формалізований, обчислюваний показник довіри. По–третє, саме питання усунення шуму – тобто вибір конкретного методу фільтрації для практичного застосування – досі не отримало кількісного обґрунтування: попередні роботи обмежувалися класифікацією шуму, не порівнюючи ефективність методів його усунення. Ця робота присвячена розв'язанню саме цих трьох завдань.

Аналіз останніх досліджень і публікацій. Алгоритм сегментації даних [1], критерії якого наведено в розділі «Матеріали та методи» для порівняння з застосованим тут підходом, показав високу відповідність ручній візуальній сегментації для стабільних і спадних сегментів; за оцінкою авторів [1], близько 5% сегментів можуть бути класифіковані помилково через різкі аномалії, не усунені на етапі попередньої обробки.

Алгоритм [2] додає очищення від аномалій та детекцію заправок за диференційованим адаптивним згладжуванням, детально описаним у розділі «Матеріали та методи»; на вибірці п'яти локомотивів (близько 0,5 млн. точок) це забезпечило повну відповідність ручній перевірці для всіх 43 виявлених заправок; частка пропущених подій у [2] не оцінювалась.

Методика класифікації шуму [3] визначає шум як залишок лінійної апроксимації сегмента і на основі його статистичних та спектральних ознак відносить сегмент до одного з типів за двома шкалами — спектральною (білий/кольоровий) і статистичною (гаусівський/дрейфовий/імпульсний); повний перелік критеріїв наведено в розділі «Матеріали та методи» (табл. 1).

Незважаючи на це, алгоритми серії побудовані послідовно — детекція заправок [2] спирається на сегментацію [1], а класифікація шуму відбувається на залишках апроксимації сегментів, — проте кожну з трьох робіт апробовано окремо, на власній вибірці й з власними параметрами, тому наскрізна робота всіх алгоритмів як єдиного конвеєра на спільних даних не перевірялась, а методика класифікації шуму [3] не супроводжувалась оцінкою ефективності методів його усунення. Це визначає актуальність об'єднання напрацювань в єдину методологію, апробація якої на ширшій вибірці даних і становить предмет цього дослідження. Ця робота не повторює результати попередніх на нових даних, а перевіряє, як змінюються їхні висновки при наскрізному виконанні алгоритмів у спільному конвеєрі.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Роботи серії виконані на різних вибірках. У цій роботі використано окрему вибірку шести локомотивів із 280 подіями заправки проти 43 у [2], тому кількісні показники різних робіт серії не є прямо порівнянними.

Мета дослідження. Метою дослідження є розробка та експериментальна апробація комплексної методології обробки телеметричних даних рівня палива маневрових локомотивів, яка об'єднує сегментацію, диференційоване згладжування, детекцію заправок, класифікацію нелінійного шуму та кількісне порівняння методів його усунення в єдиний відтворюваний конвеєр обробки з незалежною контрольною валідацією за балансовою моделлю витрати палива.

Для досягнення мети поставлено такі завдання:

- об'єднати алгоритми сегментації [1], диференційованого згладжування та детекції заправок [2] і класифікації шуму [3] в єдиний програмний конвеєр обробки даних;
- запропонувати формалізований, обчислюваний показник довіри до кожної виявленої події заправки, застосовний до вибірки такого розміру, для якого ручна верифікація кожної події є практично неможливою;
- розширити методику класифікації шуму [3] до рівня повної вибірки сегментів і отримати узагальнені статистичні характеристики шуму по всій вибірці локомотивів;
- виконати кількісне порівняння чотирьох методів усунення шуму — ковзного середнього, фільтра Калмана, вейвлет–порогової обробки та гібридного фільтра (вейвлет–порогування залишку фільтра Калмана) — за приростом відношення сигнал/шум;
- виконати незалежну валідацію конвеєра за допомогою контрольної балансової моделі погодинної витрати палива;
- задокументувати обмеження, зумовлені якістю вихідних даних, та їхній можливий вплив на достовірність отриманих результатів.

Викладення основного матеріалу дослідження. Загальна структура запропонованого конвеєра обробки даних, що об'єднує всі етапи методології — від видалення аномалій у необробленому сигналі до формування зведених результатів, — наведена на рисунку 1. Паралельні гілки класифікації шуму та порівняння методів фільтрації виконуються незалежно одна від одної після завершення диференційованого згладжування, оскільки обидві спираються на той самий ізольований шум, але не впливають одна на одну.

Дослідження виконано на телеметричних даних шести маневрових локомотивів, позначених далі як № 1 – № 6, з базовим кроком дискретизації, близьким до однієї хвилини. Кожен запис містить часову мітку, рівень палива в баку, швидкість руху та потужність дизель–генераторної установки. Обробку реалізовано у вигляді єдиного програмного конвеєра, що послідовно виконує видалення аномалій, детекцію заправок, сегментацію, класифікацію шуму за сегментами, диференційоване згладжування та порівняння методів фільтрації.

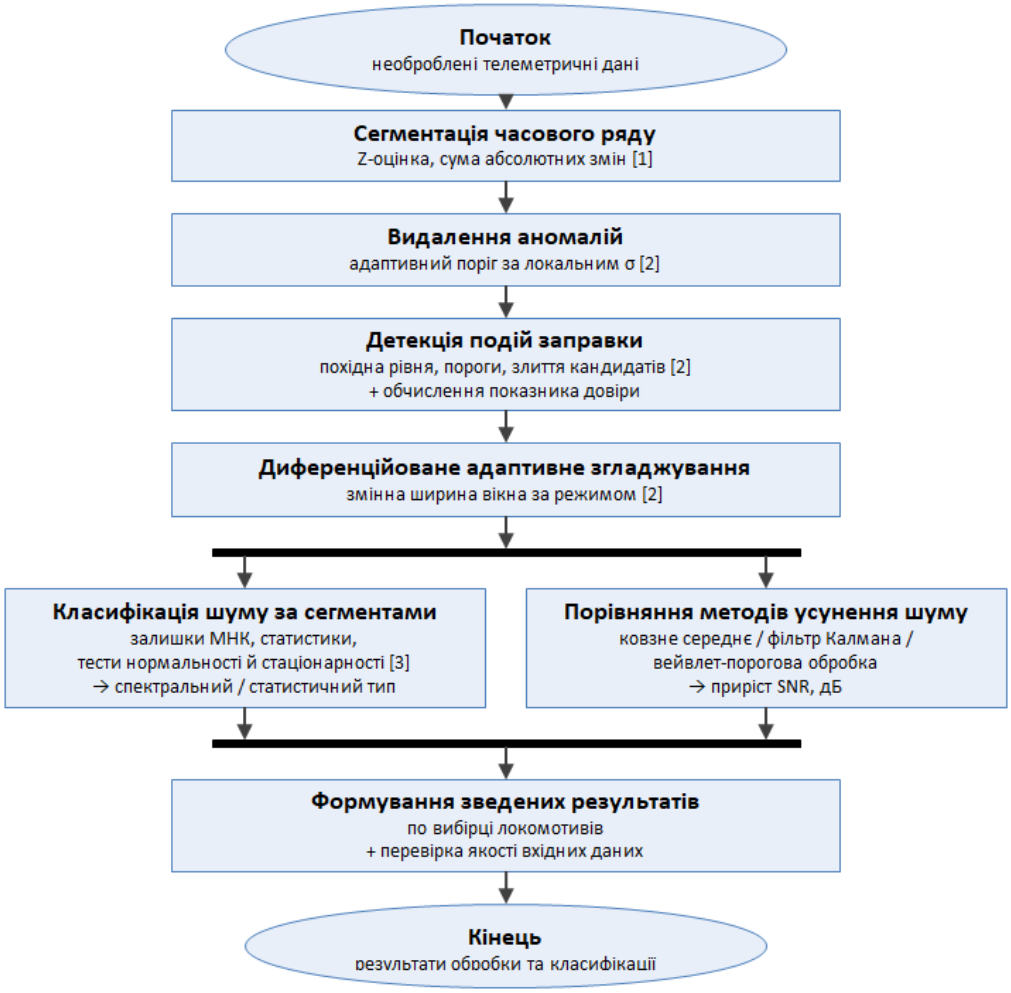


Рисунок 1 – UML–діаграма діяльності алгоритму комплексної обробки телеметричних даних

Сегментація часового ряду. Завдання етапу – виділити однорідні ділянки режиму роботи: стоянки, руху та заправки. У методиці [1] межі таких ділянок визначаються статистично: попередньо згладжений сигнал ділиться на перекриваючі вікна, для яких обчислюється Z–оцінка значущості переходу між сусідніми вікнами:

$$Z_i = \frac{|\mu_{i+1} - \mu_i|}{\sigma_i} \quad (1)$$

де Z_i – оцінка значущості переходу між вікнами i та i+1; μ_i, μ_{i+1} – середні прирости рівня палива у вікнах i та i+1; σ_i – локальне стандартне відхилення у вікні i.

Додатково для кожної потенційної межі обчислюється сума абсолютних змін даних:

$$C = \sum_{k=i}^{j-1} |x_{k+1} - x_k| \quad (2)$$

де C – сумарна зміна даних за модулем між точками i та j; x_k – значення рівня палива у точці k.

У цій роботі межі сегментів визначаються не за цими пороговими величинами, а за зміною режиму локомотива: кожній точці ряду присвоюється один із трьох типів —

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

стоянка, рух (активна потужність понад 0,1 кВт) або заправка — а межами сегментів стають точки зміни цього типу. Режим руху визначається за активною потужністю, а не за швидкістю, тобто відображає роботу дизель–генераторної установки під навантаженням; інтервали виявлених заправок мають пріоритет над режимом, присвоєним за потужністю. Порогові величини (1) і (2) запропоновано в [1]; тут вони наведені для порівняння підходів. Кожен виділений сегмент апроксимується лінійною залежністю рівня палива від часу з коефіцієнтом нахилу k_c, що визначає тип тренду сегмента.

Сусідні сегменти стоянки, розділені часовим інтервалом не більше 10 хвилин, об'єднуються, що зменшує надмірну фрагментацію даних; поріг 10 хвилин обрано за результатами контрольної валідації відновлення рівня на п'яти значеннях порога, яка показала прийнятну точність відновлення для розривів до 10 хвилин. Крім того, сегменти коротші за шість відліків приєднуються до попереднього, окрім сегментів заправки.

Диференційоване адаптивне згладжування. Після сегментації та видалення аномалій за адаптивним порогом за методикою [2] до очищеного сигналу застосовується диференційоване згладжування гаусовим фільтром зі змінною шириною вікна. Базова ширина вікна обчислюється у два кроки. Спершу визначається проміжна ширина за середньою швидкістю локомотива:

$$W_0 = \mathrm{clip}\left(60 + 120 \cdot \frac{\bar{v}}{v_н}, 41, 301\right) \quad (3)$$

де W_0 — базова ширина вікна згладжування, у відліках; $\bar{v}$ — середня швидкість локомотива за весь період спостереження; $v_н$ — нормувальна швидкість, значення якої наведено нижче; clip(x, xmin, xmax) — обмеження значення x діапазоном.

На другому кроці отримана ширина множиться на коефіцієнт, що залежить від середнього рівня палива відносно номінальної ємності бака, із повторним обмеженням тим самим діапазоном 41–301 відлік: чим повніший бак, тим ширше вікно. Нижче під базовою шириною вікна розуміється результат цього другого кроку. Числові коефіцієнти 60 і 120, нормувальна швидкість 70 км/год та межі обмеження 41–301 відліків перенесено з [2] без змін, щоб результати цього дослідження були порівнянні з попередніми. Для ділянок руху ширина вікна встановлюється втричі більшою за базову, для ділянок стоянки — у 1,5 рази більшою, що забезпечує максимальне придушення вібраційного шуму там, де точна форма сигналу не має самостійного значення. Натомість у межах кожної заправки ширина вікна мінімізується й обмежується вузьким діапазоном 3–11 відліків; конкретне значення в цих межах залежить від локального рівня палива та швидкості. Це дозволяє зберегти різкість фронту підйому рівня і точно визначити обсяг залитого палива.

Детекція подій заправки та формалізована оцінка довіри. Кандидати на заправку виділяються за стійким зростанням згладженої похідної рівня палива, яке перевищує адаптивний поріг, пропорційний локальному розкиду похідної, швидкості та потужності, з обмеженням на мінімальний обсяг підйому; кандидати малого обсягу додатково відсіюються за наявністю ознак руху в центральній точці інтервалу, тоді як підйоми великого обсягу зберігаються незалежно від цієї ознаки; сусідні у часі кандидати

об'єднуються в одну подію. На відміну від методики [2], де надійність детекції підтверджена ручною верифікацією на порівняно невеликій вибірці з 43 подій на п'яти локомотивах, у цій роботі, зважаючи на суттєво більший обсяг вибірки, запропоновано формалізований, обчислюваний показник довіри для кожної окремої події:

$$\text{Довіра} = \left(w_1 R^2 + w_2 M + w_3 I(\sigma_{\text{до}} < \theta_1) + w_4 I(\sigma_{\text{після}} < \theta_2)\right) \cdot K_{\text{руху}} \qquad (4)$$

де R^2 — коефіцієнт детермінації лінійної апроксимації рівня палива протягом заправки, що характеризує лінійність фронту підйому; M — частка неспадних кроків рівня палива в межах заправки, що характеризує монотонність; $\sigma_{\text{до}}$, $\sigma_{\text{після}}$ — локальне стандартне відхилення рівня палива в 15–хвилинних вікнах безпосередньо до і після заправки; $I(\cdot)$ — індикаторна функція, що дорівнює 1, якщо умова в дужках виконується, і 0 в іншому разі; θ_1, θ_2 — порогові значення стабільності рівня палива до і після заправки; $K_{\text{руху}}$ — штрафний коефіцієнт, що зменшує довіру, якщо середня швидкість локомотива протягом заправки перевищує допустиме значення; $w_1 - w_4$ — вагові коефіцієнти складових довіри, $w_1 + w_2 + w_3 + w_4 = 1$.

У цьому дослідженні використано значення w_1 = 0,4; w_2 = 0,3; w_3 = 0,2; w_4 = 0,1; θ_1 = 5 л; θ_2 = 8 л; $K_{\text{руху}}$ = 0,5 за середньої швидкості понад 1 км/год протягом заправки, інакше $K_{\text{руху}}$ = 1. Вибір вагових коефіцієнтів відображає відносну важливість лінійності фронту підйому як найсильнішої ознаки справжньої заправки порівняно зі стабільністю рівня до і після події. Наведені вагові коефіцієнти та порогові значення задано апріорно й на цій вибірці не калібрувались, тому абсолютне значення довіри слід трактувати як відносну міру для ранжування подій, а не як калібровану ймовірність. Калібрування цих параметрів за контрольною підвибіркою верифікації визначено як окремий напрям подальшої роботи. Ємність бака прийнято однаковою для всіх локомотивів вибірки й рівною номінальній ємності 6000 л, оскільки парк складається з локомотивів одного типу; індивідуальної паспортної ємності за кожним локомотивом у наявних даних не було. Заправки додатково класифікуються за типом за умовними межами, прийнятими для цілей звітності: повна — рівень після заправки не нижче 85% номінальної ємності; дозаправка — обсяг менше 15% номінальної ємності; часткова — в інших випадках.

Класифікація нелінійного шуму за сегментами. Для кожного сегмента режиму руху або стоянки, отриманого на етапі сегментації, за тією самою лінійною моделлю виконується окрема лінійна апроксимація методом найменших квадратів у межах сегмента, а залишки цієї апроксимації розглядаються як шум цього сегмента:

$$r_i = y_i - (\hat{a} + \hat{b} \cdot t_i) \qquad (5)$$

де r_i — залишок апроксимації, тобто шум, у точці i; y_i — фактичний рівень палива у точці i; $\hat{a}$, $\hat{b}$ — оцінки коефіцієнтів лінійної регресії (оцінені окремо для кожного сегмента); t_i — часова мітка точки i.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Для отриманого набору залишків обчислюються описові статистики — середнє, стандартне відхилення, а також вибіркові коефіцієнти асиметрії та ексцесу як нормовані третій та четвертий центральні моменти розподілу залишків у межах сегмента.

Додатково обчислюється автокореляція із зсувом 1 та спектральний нахил за методом Велча [4] як коефіцієнт нахилу лінійної регресії, оціненої методом найменших квадратів у логарифмічних координатах «логарифм спектральної щільності потужності — логарифм частоти». Для перевірки нормальності розподілу залишків застосовуються тести Андерсона–Дарлінга [9] та Колмогорова–Смірнова, для перевірки стаціонарності — розширений тест Дікі–Фуллера [7] та тест Квятковського–Філліпса–Шмідта–Шина [8].

Порогові значення, використані для розмежування підтипів спектрального шуму та для віднесення сегмента до дрейфового чи імпульсного статистичного типу, підібрані емпірично на основі попереднього дослідження [3] і не піддавалися окремому калібруванню на вибірці цієї роботи. Це свідоме методологічне рішення: застосування тих самих порогів, що й у попередній публікації, дозволяє порівнювати отримані розподіли з результатами [3] без додаткового джерела розбіжності, пов'язаного зі зміною критеріїв класифікації. Наслідки цього рішення для розподілу статистичних типів шуму розглянуто в розділі результатів.

На основі отриманих ознак кожен сегмент класифікується одночасно за двома незалежними шкалами:

Таблиця 1

Критерії класифікації шуму за сегментами

Шкала	Тип	Критерій
Спектральна	Білий	Розподіл нормальний, середнє близьке до нуля, нахил спектра близький до нуля
Спектральна	Кольоровий, з підтипами	Нахил спектра суттєво відмінний від нуля; підтип визначається величиною нахилу
Статистична	Гаусівський	Розподіл залишків не відхиляється значуще від нормального
Статистична	Дрейфовий	Значущий нормований коефіцієнт нахилу регресійної кривої залишків на час
Статистична	Імпульсний	Наявність стрибка, що суттєво перевищує стандартне відхилення
Обидві шкали	Не класифіковано	Довжина сегмента менша за 20 точок — сегмент взагалі не допускається до класифікації; довжина менша за 32 точки або нульовий розкид залишків — оцінка спектрального нахилу ненадійна, тому спектральний тип не визначається

Порівняння методів усунення шуму. Порівнюються чотири методи, застосовані до ізольованого шуму — різниці очищеного сигналу та адаптивно згладженого тренду.

Ця різниця за побудовою не містить стрибків заправок, оскільки вони вже враховані диференційованим згладжуванням.

Перший метод — фільтр ковзного середнього з фіксованим вікном шириною $W_{кс}$, що усереднює $W_{кс}$ сусідніх відліків ізольованого шуму; у цьому дослідженні $W_{кс}$ = 15 відліків, що за кроку дискретизації близько однієї хвилини відповідає вікну близько 15 хвилин — тій самій величині, що й вікна оцінки стабільності рівня до і після заправки; окремого підбору цієї ширини за критерієм якості фільтрації не виконувалось.

Другий метод — одновимірний фільтр Калмана [6] з моделлю випадкового блукання прихованого стану, прогноз і оновлення якого виконуються за канонічними рівняннями фільтра з дисперсією процесу $\sigma^2_п$ та дисперсією вимірювального шуму $\sigma^2_в$.

Дисперсія вимірювального шуму $\sigma^2_в$ оцінюється робастно — як квадрат медіанного абсолютного відхилення приростів ізольованого шуму, нормованого коефіцієнтом 0,6745, що на відміну від звичайної вибіркової дисперсії нечутливе до поодиноких залишкових викидів; дисперсія процесу приймається як $\sigma^2_п = \max(0{,}01 \cdot \sigma^2_в, 10^{-6})$, де нижня відсічка запобігає виродженню фільтра на сегментах із малим розкидом шуму. Співвідношення $\sigma^2_п/\sigma^2_в = 0{,}01$ задає повільну зміну прихованого стану порівняно з рівнем вимірювального шуму; його зафіксовано однаковим для всіх локомотивів, щоб порівняння методів не залежало від індивідуального налаштування фільтра. Саме значення 0,01 зафіксовано апріорно і на цій вибірці не калібрувалось; підбір цього відношення за критерієм якості фільтрації визначено як окремий напрям подальшої роботи.

Третій метод — вейвлет–порогова обробка. Ізольований шум розкладається багаторівневим дискретним перетворенням Хаара на одну апроксимуючу складову найгрубішого масштабу та кілька деталізуючих складових — по одній на кожен рівень розкладання, кожна зі своїм масштабом. На кожному рівні деталізуючі коефіцієнти, абсолютна величина яких не перевищує порогового значення, вважаються шумовими й зануляються, а ті, що перевищують поріг, зменшуються на його величину (м'яке порогування); далі сигнал відновлюється оберненим перетворенням. Порогове значення для кожного рівня l обчислюється за універсальним правилом Донохо–Джонстона [5]:

$$\lambda_l = \sigma_l \cdot \sqrt{2 \cdot \ln N_l} \qquad (6)$$

де λ_l — порогове значення для рівня l; σ_l — робастна оцінка стандартного відхилення деталізуючих коефіцієнтів рівня l через медіанне абсолютне відхилення; N_l — кількість деталізуючих коефіцієнтів рівня l.

Базис Хаара обрано як найпростіший ортогональний базис, що добре локалізує стрибкоподібні зміни рівня палива; використано до чотирьох рівнів розкладання. Діадичне перетворення вимагає довжини, кратної степеню двійки, тому ряд обрізається до найбільшого степеня двійки, що не перевищує його довжини, а решта відліків у кінці залишається необробленою; наслідки цього розглянуто в розділі результатів.

Четвертий метод — гібридний: до залишку фільтра Калмана, тобто різниці ізольованого шуму та калманівської оцінки, застосовується вейвлет–порогова обробка в базисі Хаара з порогом Донохо–Джонстона, але з трьома рівнями розкладання замість

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

чотирьох, після чого відфільтрований залишок додається до калманівської оцінки. Задум полягав у поєднанні здатності фільтра Калмана відслідковувати повільну складову з локалізацією стрибків у вейвлет–базисі.

Ефективність усіх чотирьох методів оцінюється єдиним показником — приростом відношення сигнал/шум:

$$\Delta\text{SNR} = 10 \cdot \log_{10}\left(\frac{\sigma_{\text{ш}}^2}{\sigma_{\text{з}}^2}\right), \text{дБ} \quad (7)$$

де $\sigma_{\text{ш}}^2$ — дисперсія ізольованого шуму до фільтрації; $\sigma_{\text{з}}^2$ — дисперсія залишку після фільтрації, тобто шум мінус відфільтрована оцінка.

Валідація балансової моделі витрати палива. Незалежно від детекції подій обчислюється контрольна балансова модель витрати палива, з двокомпонентною структурою: базова складова холостого ходу, тобто питома витрата палива за одиницю часу (л/год) за відсутності навантаження, та маргінальна складова, пропорційна активній потужності. Цю питому витрату оцінено на відліках із практично нульовою активною потужністю, маргінальну — на відліках під навантаженням, щоб уникнути змішування режимів. Обидві складові оцінюються потоко–зваженим способом у межах відповідного режиму на адаптивно згладженому сигналі з вилученням вікон навколо заправок і різких від'ємних стрибків рівня. Потоко–зважене оцінювання — відношення сумарної витрати пального до сумарного часу режиму, а не середнє від покрокових приростів — застосовано через квантування телеметрії: рівень палива передається з обмеженою роздільною здатністю датчика, тому на холостому ходу приріст рівня за окремий крок часто дорівнює нулю, і покрокові оцінки витрати систематично зміщені. Відношення сум зберігає баланс маси незалежно від розподілу кроків. Отримані погодинні прогнози зіставляються з фактичною погодинною витратою на «чистих» годинах, тобто тих, де не менше 75% часу лежить поза вікнами подій і наявно не менше півгодини валідних відліків. Якість узгодження оцінюється квадратом коефіцієнта кореляції Пірсона між прогнозом і фактом, далі позначеним R^2; така міра відображає тісноту лінійного зв'язку, але не враховує систематичного зміщення прогнозу, тому кількісно не тотожна класичному коефіцієнту детермінації.

Результати та їх обговорення. Оброблено дані шести локомотивів; виявлено 280 подій заправки. Придушення шуму на етапі адаптивного згладжування, оцінене незалежно на сегментах стоянки як відношення робастних оцінок розкиду залишків до й після обробки, становить у середньому 1,9 дБ із діапазоном за локомотивами від −0,01 до 6,1 дБ; від'ємне значення для одного локомотива означає, що на його сегментах стоянки згладжування не дало вимірного придушення шуму. Статус надійності визначено формально: позначку «потребує перевірки датчика» присвоєно локомотиву, для якого R^2 балансової моделі нижчий за 0,1 за кількості виявлених заправок менше п'яти. Загальні показники за локомотивами наведено в табл. 2.

Таблиця 2

Загальні показники за локомотивами вибірки

№ локомотиву	Заправок	R^2 балансової моделі	Статус надійності
1	23	0,194	Задовільний
2	8	0,521	Задовільний
3	118	0,433	Задовільний
4	46	0,319	Задовільний
5	82	0,357	Задовільний
6	3	0,032	Потребує перевірки датчика

Детекція подій заправки. Із 280 виявлених заправок повними класифіковано 244 події, або 87,1%, частковими — 31 подію, або 11,1%, дозаправками — 5 подій, або 1,8%. Середній обсяг заправки становить 2910 л, медіана — 2997 л, діапазон — від 356 до 5275 л, а сумарний обсяг за весь період спостереження — 814 838 л. Розподіл обсягів заправок близький до одномодального з лівосторонньою (від'ємною) асиметрією: середнє значення (2910 л) нижче за медіану (2997 л), що загалом відповідає характеру, описаному в роботі [2] для меншої вибірки.

Середня довіра детекції становить 0,828, медіана — 0,863; довіру не нижче 0,7 мають 234 з 280 подій, або 83,6%, а не нижче 0,9 — 112 подій, або 40,0%. Такий розподіл довіри дозволяє зосередити ручну перевірку на меншості подій із низькою довірою: події з довірою нижчою 0,7, а це 46 подій, або 16,4% вибірки, варто розглядати як кандидатів, що потребують додаткової уваги, зокрема через можливий рух локомотива протягом заправки або нестабільність рівня палива до чи після події.

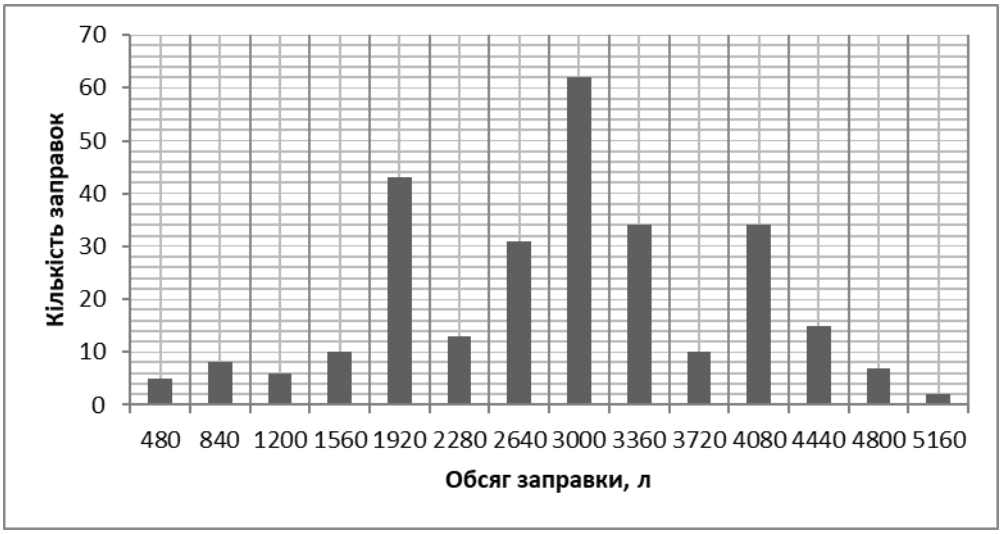


Рисунок 2 – Розподіл обсягів виявлених заправок

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

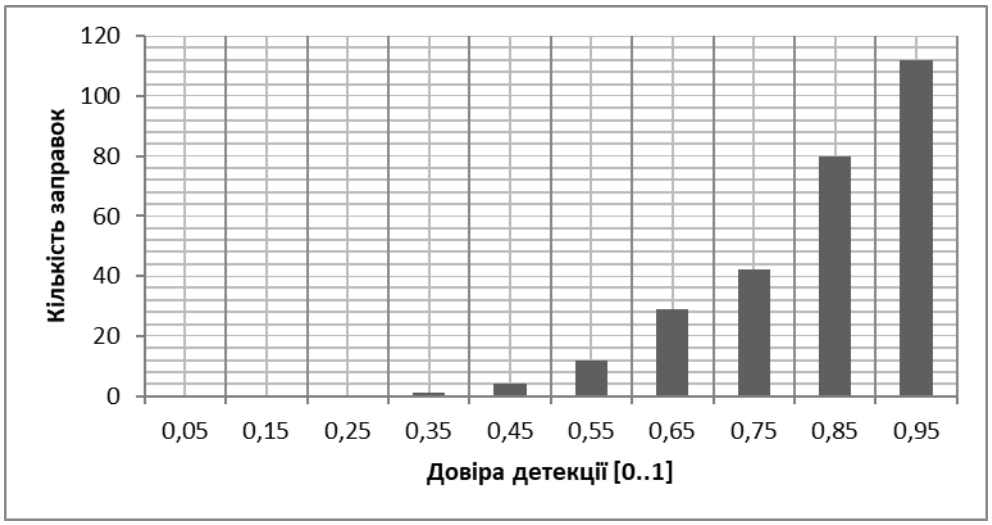


Рисунок 3 – Розподіл довіри детекції заправок

Розподіл значень довіри має виражений лівий хвіст: приблизно кожна шоста подія, а це 16,4%, отримує довіру нижче 0,7, переважно через нестабільність рівня палива до або після події: із 46 таких подій у 39 локальне стандартне відхилення рівня до заправки перевищує поріг 5 л, а у 33 — відхилення після заправки перевищує поріг 8 л. Ще однією причиною є недостатня лінійність фронту підйому рівня, що може вказувати або на повільну, нерівномірну заправку, або на накладання шуму на короткому інтервалі: у 13 із цих 46 подій коефіцієнт детермінації фронту нижчий за 0,5. Штрафний коефіцієнт $K_{\text{руху}}$ у цій вибірці не активувався жодного разу: всі 280 виявлених заправок відбулися на стоянці, тому R^2 на розподіл довіри в цій вибірці він не вплинув.

Порівняно з роботою [2], де підтверджено 43 заправки на вибірці 5 локомотивів, отримана в цьому дослідженні вибірка 280 подій на 6 локомотивах якісно інша за складом: вона містить усі виявлені конвеєром події без попереднього відбору, зокрема й події з нижчою довірою, які при ручній верифікації меншої вибірки [2] могли бути відсіяні. Тому пряме зіставлення часток «надійних» подій у двох роботах некоректне.

Класифікація нелінійного шуму. Усього виділено 67 575 сегментів режимів руху та стоянки, із яких до класифікації шуму допущено 27 792 сегменти, довжина яких становить не менше 20 відліків. За цими сегментами спектральний тип шуму розподілено так: кольоровий шум становить 53,5%, білий — 1,6%, а для решти, а це 44,9%, спектральний нахил оцінити не вдалось, тому спектральний тип не визначався: для найкоротших із допущених сегментів оцінка спектральної щільності за методом Велча [4] стає ненадійною через властивий цьому методу компроміс між зміщенням і роздільною здатністю на коротких реалізаціях. Для коректного зіставлення з [3], де класифіковано лише достатньо довгі сегменти, частки слід перераховувати на базу класифікованих сегментів: кольоровий шум становить 97,1%, білий — 2,9%, що близько до 99,77% і 0,23%, отриманих у [3], де практично весь кольоровий шум мав коричнево–подібний характер. Серед підтипів кольорового шуму переважають ті, що наближаються до коричневого, 26,9% усіх сегментів, або 50,3% кольорових, та до рожевого, 25,4%

усіх сегментів, або 47,5% кольорових, що узгоджується з переважно від'ємним медіанним нахилом спектра.

За статистичним типом переважає гаусівський шум із часткою 79,0%, змішаний ненормальний тип — залишкова категорія, яка не виділяється окремо в табл. 1 і охоплює сегменти, залишки яких значуще відхиляються від нормального розподілу, але не відповідають ні критерію дрейфового, ні критерію імпульсного типу, — становить 15,2%, імпульсний — 5,8%. Дрейфовий статистичний тип у цій вибірці не виявлено жодного разу, що протилежно результату [3], де цей тип домінував; можливі причини розглянуто нижче.

Мінімальну довжину сегмента прийнято рівною 20 точкам: на коротших вибірках потужність тесту Андерсона–Дарлінга [9] істотно знижується. Через це до класифікації не допущено 39 783 із 67 575 сегментів, або 58,9%, що свідчить про короткочасність значної частини режимів роботи і вказує на потребу або знизити цей поріг із корекцією критеріїв, або об'єднувати сусідні короткі сегменти одного режиму перед класифікацією.

Повна відсутність дрейфового статистичного типу є найсуттєвішим результатом цього розділу. У [3] дрейфовий тип домінував (94,33% проти 1,67% гаусівського та 4% імпульсного); його зникнення потребує окремого розгляду. Як і в [3], шумом тут вважаються залишки лінійної апроксимації сегмента, і класифікація виконується до етапу диференційованого згладжування — отже розбіжність не можна пояснити вилученням трендових складових згладжуванням. Натомість відрізняються дві методичні обставини. Перша — спосіб виділення сегментів: тут їх межі визначаються за зміною режиму локомотива, а класифікується ряд після видалення аномалій із заповненням викинутих відліків, тобто сегменти однорідніші за режимом і не містять викидів, які могли б давати значущий нахил залишків. Друга — перенесене без змін з [3] порогове значення нормованого нахилу регресії залишків, яке для сегментів цієї вибірки може бути надто консервативним. Не виключено також, що дрейфова складова шуму менш характерна для локомотивів цієї конкретної вибірки. Розмежування цих трьох причин — калібрування порога на цій вибірці та класифікація шуму на тих самих сегментах до і після згладжування — визначено як окремий напрям подальшої роботи.

Порівняння методів усунення шуму. Фільтр ковзного середнього забезпечив приріст відношення сигнал/шум від 1,8 до 4,2 дБ, у середньому по вибірці 2,9 дБ, фільтр Калмана з робастною оцінкою параметрів – від 1,3 до 2,0 дБ, у середньому 1,7 дБ. Гібридний фільтр дав практично той самий результат, що й фільтр Калмана, – від 1,3 до 2,0 дБ, у середньому 1,7 дБ, при чому дисперсія його залишку відрізняється від калманівської менш ніж на 1% для всіх шести локомотивів і не наближається до вейвлетової. Це означає, що вейвлет–порогування, застосоване до залишку фільтра Калмана, майже не змінює результату: калманівський етап уже видаляє високочастотну складову, і поріг Донохо–Джонстона на такому залишку спрацьовує слабо. Аномально високий приріст відношення сигналу/шуму дає лише вейвлет–порогова обробка, застосована безпосередньо до ізольованого шуму (третій метод, 16,7–27,5 дБ), тоді як гібри-

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

дний фільтр із тим самим порогуванням у базисі Хаара залишається на рівні фільтра Калмана (1,3–2,0 дБ). Отже, причина завищення – у застосуванні м'якого порогування безпосередньо до нефільтрованого шуму, а не у властивостях вейвлет–базису. Обидва результати виглядають узгодженими між собою за порядком величини: обидва методи залишають у залишку від 38% до 74% вхідної дисперсії шуму (див. нижче), що відповідає отриманим приростам одного порядку величини для шуму з додатною автокореляцією сусідніх відліків. Ці величини не зіставляються з приростом 32,6 дБ, наведеним у [2]: там йдеться про приріст на етапі адаптивного згладжування необробленого сигналу, тоді як тут методи застосовано до вже ізольованого шуму, з якого трендову складову вилучено на попередньому етапі.

Вейвлет–порогова обробка формально показала приріст від 16,7 до 27,5 дБ, у середньому 22,3 дБ, що на порядок перевищує результати двох інших методів. Причина цієї розбіжності – властивість самого показника в застосуванні до ізольованого шуму: він порівнює дисперсію вхідного шуму з дисперсією різниці між шумом і виходом фільтра, тому найвищу оцінку отримує той метод, який найменше змінює вхідний сигнал. Дисперсія залишку для вейвлет–порогової обробки становить від 0,2% до 2,1% вхідної дисперсії шуму, тоді як для ковзного середнього – від 38% до 67%, а для фільтра Калмана – від 63% до 74%. Це означає, що м'яке порогування залишає шум практично незмінним, і високий приріст відображає саме це, а не кращу якість фільтрації. Отже, за наявним показником вейвлет–порогову обробку не можна визнати кращим методом, а сам приріст відношення сигнал/шум не розрізняє якісну фільтрацію й підлаштування під вхідний шум. Для коректного порівняння потрібна метрика, яка враховує спотворення корисної складової. На цей самий висновок вказує й формальний критерій, закладений у сам конвеєр: ознаку підозри на підлаштування під вхідний шум виставлено для всіх шести локомотивів вибірки без винятку, тобто аномально високий приріст вейвлет–порогової обробки є системним для всього парку, а не особливістю окремих локомотивів.

Таблиця 3

Порівняння чотирьох методів усунення шуму за приростом відношення сигнал/шум

Метод	Приріст SNR, дБ (діапазон)	Приріст SNR, дБ (середнє)	Застосовність
Ковзне середнє, вікно 15	1,8–4,2	2,9	Задовільна
Фільтр Калмана з робастними σ^2в і σ^2п	1,3–2,0	1,7	Задовільна
Вейвлет–порогова обробка	16,7–27,5	22,3	Не оцінюється
Гібридний фільтр	1,3–2,0	1,7	Задовільна

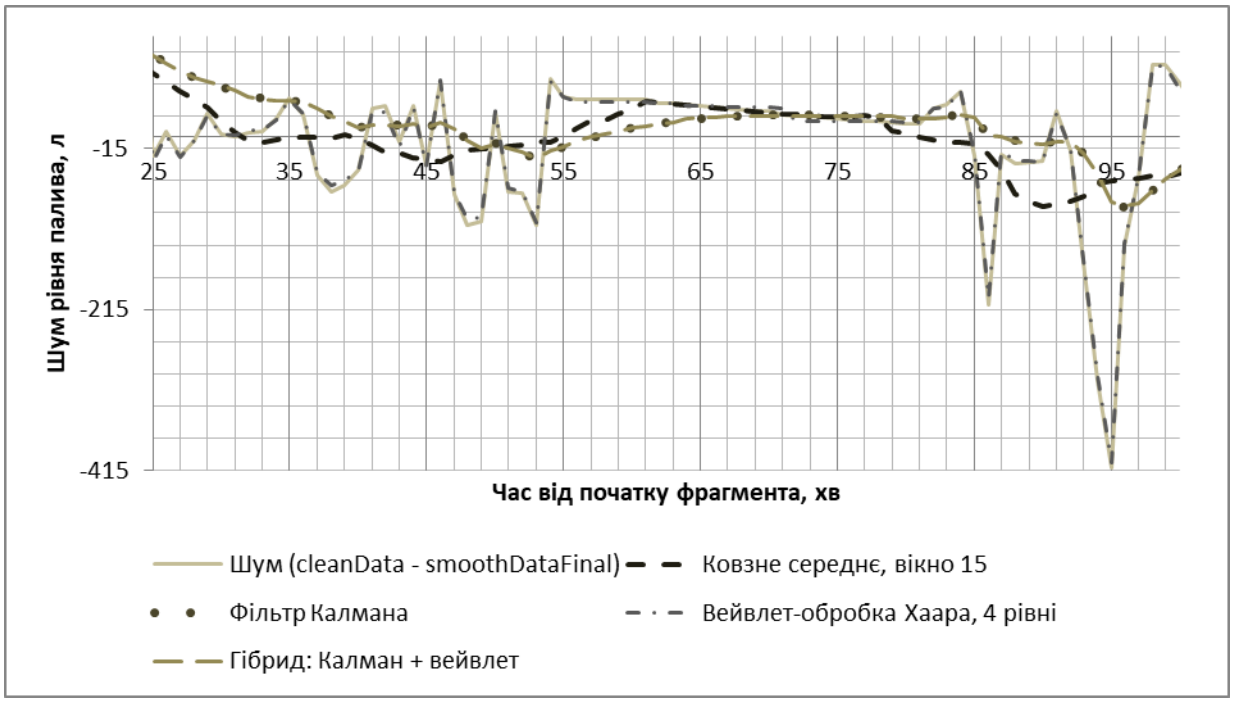


Рисунок 4 – Результат застосування методів усунення шуму до фрагмента ізольованого шуму

Помірна, але стабільна перевага ковзного середнього над фільтром Калмана у середньому на 1,2 дБ обумовлена невідповідністю моделі: оптимальність фільтра Калмана доведена для лінійно–гаусівської моделі з коректно заданими дисперсіями [6], тоді як тут модель процесу свідомо спрощена до випадкового блукання, а реальний шум кольоровий. За такої невідповідності гарантія оптимальності не діє, і простіший метод може виявитися точнішим.

На завищену оцінку вейвлет–порогової обробки, крім властивостей самого показника, діють два методичні чинники. Поріг Донохо–Джонстона [5] обґрунтований для незалежного гаусівського шуму, тоді як переважна частина шуму цієї вибірки кольорова, тому порогування виявилося надто слабким, а не надмірно агресивним — що й узгоджується з надзвичайно низькою дисперсією залишку (0,2–2,1% вхідної). Другий чинник — обрізання ряду: необроблений «хвіст» відліків входить у розрахунок залишку з нульовою похибкою і додатково занижує його дисперсію. Розмежування цих чинників потребує перевірки на синтетичному сигналі з відомим рівнем шуму.

Балансова модель витрати палива демонструє помірну лінійну узгодженість із фактичною погодинною витратою, R^2 = 0,19–0,52, для п'яти з шести локомотивів вибірки (табл. 2). Такі значення відповідають лише частковому узгодженню моделі з даними; за наявної кількості спостережень на локомотив оцінки R^2 спираються на обмежену кількість «чистих» годин, тому їх слід трактувати як орієнтовні, а не як підтвердження адекватності моделі. Єдиний виняток — локомотив № 6, для якого R^2 моделі становить лише 0,032. Цей локомотив вирізняється мінімальною кількістю заправок — три — та переважно холостим режимом роботи, що звужує діапазон варіації погодинної витрати

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

й ускладнює оцінку якості моделі незалежно від точності датчика. Оскільки за наявними даними розрізнити низьку варіативність режиму й несправність датчика неможливо, для цього локомотива рекомендовано окрему перевірку показань датчика рівня. Розподіл R^2 за локомотивами наведено в табл. 2. На результати класифікації шуму та детекції заправок це обмеження не поширюється: обидва етапи спираються на локальні характеристики сигналу, а не на інтегральний баланс витрати за період.

Порогові значення класифікатора шуму та параметри вейвлет–порогової обробки перенесено з попередніх робіт без калібрування на цій вибірці; відсутність дрейфового типу та аномально високий приріст відношення сигнал/шум для вейвлет–порогової обробки свідчать про доцільність такого калібрування на розширеній або незалежній вибірці даних.

Запропонований формалізований показник довіри детекції заправки може бути безпосередньо використаний диспетчерськими та обліковими службами як фільтр для автоматичного пріоритезування подій, що потребують ручної перевірки: замість перегляду всіх 280 подій достатньо перевірити 46 подій із довірою нижче 0,7.

Результати класифікації шуму означають, що застосування методів, розроблених у припущенні білого шуму (зокрема й самої вейвлет–порогової обробки з універсальним порогом у поточній реалізації) варто розглядати з обережністю. Натомість помірний і передбачуваний приріст SNR від ковзного середнього й фільтра Калмана робить ці два методи обґрунтованим вибором для впровадження.

Прозоро задокументоване обмеження якості даних для одного локомотива дозволяє спрямувати ресурси на перевірку конкретного датчика замість повторної перевірки всього парку.

Висновки. У результаті дослідження запропоновано комплексну методологію обробки телеметричних даних рівня палива маневрових локомотивів, яка об'єднує в єдиний конвеєр раніше опубліковані алгоритми сегментації [1] та детекції заправок з диференційованим згладжуванням [2], розширює методику класифікації шуму [3] до рівня повної вибірки сегментів і додає новий елемент — кількісне порівняння чотирьох методів усунення шуму. Застосування методики класифікації [3] до повної вибірки сегментів дало розподіл статистичних типів, який суттєво відрізняється від отриманого в [3] на меншій вибірці: дрейфовий тип, що там домінував, тут не виявлено жодного разу. Оскільки класифікація в обох роботах виконується до згладжування, розбіжність слід відносити до способу виділення сегментів і до некаліброваного порога дрейфового типу, а не до етапу обробки; розмежування цих причин визначено як напрям подальших досліджень.

Запропонований формалізований показник довіри детекції дозволяє пріоритезувати події для ручної перевірки на вибірках, де повна верифікація практично неможлива: перевірки потребує лише меншість подій із низькою довірою. Прямого порівняння з точністю, заявленою в [2], не проводилося.

З чотирьох порівняних методів до впровадження придатні два — фільтр ковзного середнього та фільтр Калмана з робастною оцінкою параметрів, які дають помірний і

передбачуваний приріст відношення сигнал/шум. Головний методичний результат цього порівняння — не рейтинг фільтрів, а виявлена непридатність самого показника: приріст відношення сигнал/шум, обчислений на ізольованому шумі, винагороджує фільтр, який найменше змінює вхідний сигнал, і тому не розрізняє якісну фільтрацію й підлаштування під шум. За цим критерієм вейвлет–порогову обробку не можна ні визнати кращою, ні рекомендувати до впровадження.

Незалежна валідація за контрольною балансовою моделлю показала лише часткове узгодження з даними, тому конвеєр слід вважати апробованим на рівні локальних характеристик сигналу — сегментації, детекції заправок і класифікації шуму, — але не на рівні інтегрального балансу витрати. Окремо задокументовано обмеження якості вхідних даних для одного локомотива вибірки, що важливо для коректної інтерпретації результатів у подальших дослідженнях.

Перспективними напрямами подальших досліджень є: запровадження метрики порівняння фільтрів, яка враховує спотворення корисної складової, зокрема перевірка на синтетичному сигналі з відомим рівнем шуму, та обробка всього ряду з доповненням його до потрібної довжини замість обрізання; уточнення критерію дрейфового статистичного типу шуму; порівняння формалізованого показника довіри детекції заправок із результатами ручної верифікації на контрольній підвибірці; адаптація розробленого конвеєра обробки для роботи в режимі, наближеному до реального часу.

ЛІТЕРАТУРА

1. Іващев Д., Герасимов В. Алгоритм сегментації даних вимірювань рівня палива на транспорті. Вимірювальна та обчислювальна техніка в технологічних процесах. 2025. № 2 (82). С. 419–425. DOI: 10.31891/2219–9365–2025–82–60.

2. Іващев Д., Герасимов В. Виділення подій заправки та обробка даних рівня палива маневрових локомотивів. Вимірювальна та обчислювальна техніка в технологічних процесах. 2025. № 4 (84). С. 323–326. DOI: 10.31891/2219–9365–2025–84–37.

3. Іващев Д., Герасимов В. Аналіз шуму в даних вимірювання рівня палива на транспорті. Вимірювальна та обчислювальна техніка в технологічних процесах. 2025. № 3 (83). С. 385–392. DOI: 10.31891/2219–9365–2025–83–47.

4. Welch P. D. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Transactions on Audio and Electroacoustics. 1967. Vol. 15, No. 2. P. 70–73.
DOI: 10.1109/TAU.1967.1161901.

5. Donoho D. L., Johnstone I. M. Ideal spatial adaptation by wavelet shrinkage. Biometrika. 1994. Vol. 81, No. 3. P. 425–455. DOI: 10.1093/biomet/81.3.425.

6. Kalman R. E. A new approach to linear filtering and prediction problems. Journal of Basic Engineering. 1960. Vol. 82, No. 1. P. 35–45. DOI: 10.1115/1.3662552.

7. Dickey D. A., Fuller W. A. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association. 1979. Vol. 74, No. 366a. P. 427–431. DOI: 10.1080/01621459.1979.10482531.

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

8. Kwiatkowski D., Phillips P. C. B., Schmidt P., Shin Y. Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics. 1992. Vol. 54, No. 1–3. P. 159–178. DOI: 10.1016/0304–4076(92)90104–Y.
9. Anderson T. W., Darling D. A. A test of goodness of fit. Journal of the American Statistical Association. 1954. Vol. 49, No. 268. P. 765–769. DOI: 10.1080/01621459.1954.10501232.

REFERENCES

1. Ivashchev, D., & Herasymov, V. (2025). Alhorytm sehmentatsii danykh vymiriuvan rivnia palyva na transporti [Algorithm for segmentation of fuel level measurement data in transport]. Vymiriuvalna ta Obchysliuvalna Tekhnika v Tekhnolohichnykh Protsesakh, 2(82), 419–425. https://doi.org/10.31891/2219–9365–2025–82–60
2. Ivashchev, D., & Herasymov, V. (2025). Vydilennia podii zapravky ta obrobka danykh rivnia palyva manevrovykh lokomotyviv [Refueling event detection and processing of fuel level data of shunting locomotives]. Vymiriuvalna ta Obchysliuvalna Tekhnika v Tekhnolohichnykh Protsesakh, 4(84), 323–326. https://doi.org/10.31891/2219–9365–2025–84–37
3. Ivashchev, D., & Herasymov, V. (2025). Analiz shumu v danykh vymiriuvannia rivnia palyva na transporti [Noise analysis in transport fuel level measurement data]. Vymiriuvalna ta Obchysliuvalna Tekhnika v Tekhnolohichnykh Protsesakh, 3(83), 385–392. https://doi.org/10.31891/2219–9365–2025–83–47
4. Welch, P. D. (1967). The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. IEEE Transactions on Audio and Electroacoustics, 15(2), 70–73. https://doi.org/10.1109/TAU.1967.1161901
5. Donoho, D. L., & Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. Biometrika, 81(3), 425–455. https://doi.org/10.1093/biomet/81.3.425
6. Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Journal of Basic Engineering, 82(1), 35–45. https://doi.org/10.1115/1.3662552
7. Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366a), 427–431. https://doi.org/10.1080/01621459.1979.10482531
8. Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics, 54(1–3), 159–178. https://doi.org/10.1016/0304–4076(92)90104–Y
9. Anderson, T. W., & Darling, D. A. (1954). A test of goodness of fit. Journal of the American Statistical Association, 49(268), 765–769.
https://doi.org/10.1080/01621459.1954.10501232

Received 15.06.2026.
Accepted 17.08.2026.
Published 30.10.2026

A comprehensive methodology for processing fuel–level telemetry dat of shunting locomotives: nonlinear noise analysis and refueling event detection

Analysis of recent research and publications. The authors' previous work proposed an algorithm for time–series segmentation [1], differentiated adaptive smoothing with refueling detection [2], and classification of nonlinear noise by spectral and statistical type [3]. Each was validated on its own sample, and the noise classification is not accompanied by any as-

sessment of noise–removal effectiveness, which motivates merging these results into a single methodology.

Purpose of the study. To combine three previously published algorithms into a single fuel–level telemetry processing pipeline and to validate their joint operation on a common sample, with independent validation via a fuel–balance model.

Presentation of the main material. Three previously published algorithms were combined into a single pipeline, with a formalized confidence score added for every detected refueling event, accounting for the linearity and monotonicity of the level rise, level stability around the event, and a movement penalty. Noise classification was extended to the full segment sample, and four noise–removal methods were assessed by signal–to–noise gain. Validation used six locomotives: 280 refueling events at a mean confidence of 0.83, and 27,792 segments classified.

Conclusions. Colored noise predominates over white and the Gaussian type over the rest; the drift type that dominated in [3] was not observed in this sample. The moving average and Kalman filter give a stable gain of 1.3–4.2 dB, and the hybrid filter stays within 1.3–2.0 dB, whereas the formally highest wavelet figure reflects a property of the metric, which rewards the smallest change to the input noise. The balance model shows moderate linear agreement with the actual hourly consumption (R^2 = 0.19–0.52) for five of six locomotives — a control rather than a predictive tool.

Keywords: time series processing, digital signal filtering, adaptive differential smoothing, wavelet thresholding, robust estimation, event detection, noise classification, signal–to–noise ratio.

Іващев Даніїл Вадимович – аспірант кафедри електронних обчислювальних машин, Дніпровський національний університет імені Олеся Гончара, Дніпро, Україна.
ORCID https://orcid.org/0009–0002–0556–9418
Герасимов Володимир Володимирович – кандидат технічних наук, доцент, завідувач кафедри комп'ютерних наук та інформаційних технологій, Дніпровський національний університет імені Олеся Гончара, Дніпро, Україна.
ORCID https://orcid.org/0000–0002–1366–715X

Ivashchev Daniil Vadymovych – PhD student, Department of Electronic Computing Machinery, Oles Honchar Dnipro National University, Dnipro, Ukraine.
ORCID https://orcid.org/0009–0002–0556–9418
Herasymov Volodymyr Volodymyrovych – PhD (Tech), Associate Professor, Head of the Department of Computer Science and Information Technologies, Oles Honchar Dnipro National University, Dnipro, Ukraine.
ORCID https://orcid.org/0000–0002–1366–715X

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)