

<https://doi.org/10.31891/2219-9365-2026-87-39>

УДК 681.128+681.518.2

ІВАЩЕВ Данііл

Дніпровський національний університет імені Олеся Гончара

<https://orcid.org/0009-0002-0556-9418>

ivashchev_d@365.dnu.edu.ua

ГЕРАСИМОВ Володимир

Дніпровський національний університет імені Олеся Гончара

<https://orcid.org/0000-0002-1366-715X>

herasymov_v@365.dnu.edu.ua

КОРОЛЬ Анатолій

Мукачівський державний університет

<https://orcid.org/0009-0009-0585-5706>

antkorl160@gmail.com

СТАТИСТИКА СЕГМЕНТІВ РІВНЯ ПАЛИВА З ІНТЕРПОЛЯЦІЄЮ ПРОПУСКІВ

У статті представлено результати комплексного статистичного аналізу телеметричних даних маневрових тепловозів, отриманих бортовими системами моніторингу. Об'єктом дослідження є часові ряди рівня палива, швидкості руху та потужності дизель-генераторної установки, предметом — методи відновлення цілісності таких рядів і їх подальшої сегментації. Роботу зосереджено на обробці вимірювальних даних. Розглянуто проблему відновлення пропущених значень у часових рядах за допомогою методів лінійної інтерполяції. Проведено загальний аналіз суцільного масиву даних обсягом 4 898 881 хвилинних вимірювань за парком із шести локомотивів. Для контролю нефізичних стрибків сигналу спільно застосовано дискретну швидкість зміни рівня палива та фільтр із гаусовим ваговим вікном. Окремо встановлено, що 97 494 відліки лежать нижче відомого порогового рівня заповнення бака; такі відліки замінені інтерпольованими значеннями. Встановлено, що інтерполяція мікророзривів тривалістю до 10 хвилин, а таких випадків 325, відновлює неперервність ряду рівня палива, тоді як для 53 макророзривів із максимальною тривалістю до 195 діб наскрізна інтерполяція не застосовується. Визначено статистичні характеристики розподілу рівня палива: коефіцієнт асиметрії ряду становить від -0,510 до -0,247, ексцес — від 2,14 до 4,35. Після відновлення цілісності ряду масив розділено на 67 855 експлуатаційних сегментів, серед яких ідентифіковано 280 заправок. Мікророзриви становлять лише 0,022 % масиву. Контрольними експериментами показано, що відновлення значно більшої частки відліків — 2,0 % масиву, точок нижче відомого порогового рівня заповнення бака змінює кількість сегментів на 0,015 % і заправок — лише на одну; гіпотетичне пряме відновлення саме мікророзривів (0,022 % масиву) змінює кількість сегментів на 0,079 % і заправок — також лише на одну.

Ключові слова: телеметрія; рівень палива; часовий ряд; лінійна інтерполяція; згладжування гаусовим вікном; сегментація даних; якість вимірювальних даних; маневровий тепловоз.

IVASHCHEV Daniil, HERASYMOV Volodymyr

Oles Honchar Dnipro National University

KOROL Anatolii

Mukachevo State University

STATISTICS OF FUEL LEVEL SEGMENTS WITH INTERPOLATION OF GAPS

The paper presents the results of a comprehensive statistical analysis of telemetry data collected from shunting diesel locomotives by on-board monitoring systems. The object of the study is the time series of fuel level, running speed and diesel-generator set power; the subject is the set of methods for restoring the integrity of such series and for their subsequent segmentation. The work focuses on the processing of measurement data. The problem of recovering missing values in time series by means of linear interpolation is considered. A continuous data array of 4,898,881 minute-by-minute measurements collected from a fleet of six locomotives is analyzed. The discrete rate of change of the fuel level and a Gaussian-weighted moving-average filter are applied jointly to control non-physical signal jumps. Separately, 97,494 samples were found to lie below the known threshold fill level of the tank; these samples were replaced with interpolated values. It is established that interpolation of micro-gaps lasting up to 10 minutes, of which 325 were recorded, restores the continuity of the fuel level series, whereas for 53 macro-gaps with a maximum recorded duration of 195 days end-to-end interpolation is not applied. The statistical characteristics of the fuel level distribution are determined: the skewness of the series ranges from -0.510 to -0.247 and the kurtosis from 2.14 to 4.35. After the integrity of the series was restored, the array was divided into 67,855 operational segments, among which 280 refuellings were identified. Micro-gaps account for only 0.022 % of the array. Control experiments showed that recovering a much larger share of samples (2.0 % of the array — points below the known threshold fill level of the tank) changes the number of segments by 0.015 % and the number of refuellings by only one; a hypothetical direct recovery of the micro-gaps themselves (0.022 % of the array) changes the number of segments by 0.079 % and the number of refuellings likewise by only one.

Keywords: telemetry; fuel level; time series; linear interpolation; Gaussian-weighted smoothing; data segmentation; measurement data quality; shunting locomotive.

Стаття надійшла до редакції / Received 03.06.2026

Прийнята до друку / Accepted 31.07.2026

Опубліковано / Published 10.09.2026



This is an Open Access article distributed under the terms of the [Creative Commons CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

© ІВАЩЕВ Данііл, ГЕРАСИМОВ Володимир, КОРОЛЬ Анатолій

ПОСТАНОВКА ПРОБЛЕМИ У ЗАГАЛЬНОМУ ВИГЛЯДІ ТА ЇЇ ЗВ'ЯЗОК ІЗ ВАЖЛИВИМИ НАУКОВИМИ ЧИ ПРАКТИЧНИМИ ЗАВДАННЯМИ

В умовах експлуатації маневрових тепловозів системи телеметричного моніторингу є одним з інструментів контролю витрати палива та технічного стану паливної системи [1]. Разом з тим особливості вимірювального каналу рівня палива [2] та коливання рідини в баках під час руху [3] ускладнюють отримання неперервного та однорідного ряду. Окремою проблемою є втрата цілісності часового ряду через пропуски даних — відліки, відсутні у часовій сітці. Пропуски порушують рівномірність часової сітки: на межі розриву обчислення швидкості зміни рівня палива дає нефізичне значення, а межа експлуатаційного сегменту може бути визначена хибно.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Дослідження, дотичні до цієї роботи, охоплюють кілька напрямів: фізику вимірювального каналу рівня палива, методи відновлення часових рядів та алгоритми їх сегментації.

Огляд типів датчиків рівня палива — резистивних, ємнісних та ультразвукових — наведено в [2]. З цього огляду видно, що ультразвукові датчики не мають рухомих частин і безпосереднього контакту з паливом, тому придатні для тривалої експлуатації на тяговому рухомому складі, проте чутливі до плескання палива під час руху та до температурної залежності швидкості звуку й мають нижню межу вимірювання, нижче якої показання недостовірні. На тепловозах, дані яких були використані для дослідження, встановлено саме ультразвукові датчики рівня палива [4]. Саме ці особливості вимірювального каналу визначають дві вади вихідного масиву, яким присвячена ця робота: високочастотну шумову складову ряду рівня палива із середньоквадратичним відхиленням від 4,8 до 27,1 л за об'єктами та показання нижче робочого діапазону датчика, які надалі трактуються як пропуски й підлягають інтерполяції.

Проблему відновлення пропущених значень у даних промислових датчиків із великими розривами системно опрацьовано в [5], де показано, що для тривалих пропусків прості інтерполяційні схеми дають нефізичні результати. Цей висновок узятो за орієнтир для самої ідеї обмежити наскрізну інтерполяцію границею тривалості розриву, тоді як конкретне значення цієї межі визначено в цій роботі експериментально. Алгоритмічні засади сегментації рядів рівня палива на транспорті, використані далі для виділення станів руху, стоянки та заправки, наведено в попередній роботі авторів [6].

Водночас наведені джерела не дають прямої відповіді на два практичні питання, що постають під час обробки телеметрії такого типу: якою є гранична тривалість розриву, за якої лінійна інтерполяція ще залишається допустимою на реальних експлуатаційних даних, і наскільки відновлення пропусків змінює результат подальшої сегментації. Саме ці два питання й розглянуто в цій роботі на конкретному експлуатаційному масиві.

ФОРМУЛЮВАННЯ ЦІЛЕЙ СТАТТІ

Метою дослідження є статистичний аналіз телеметричних масивів рівня палива з відновленням пропусків та подальшою сегментацією за експлуатаційними станами. Для її досягнення поставлено такі завдання:

- 1) розділити причини появи пропусків і встановити їх розподіл за тривалістю;
- 2) кількісно обґрунтувати граничну тривалість розриву, за якої лінійна інтерполяція залишається допустимою, та порівняти її з альтернативними схемами відновлення;
- 3) оцінити вплив відновлення пропусків на результат сегментації та детекції заправок;
- 4) визначити показники якості попередньої обробки сигналу.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

Методологія та математичний апарат

Загальний телеметричний потік формалізується у вигляді дискретного векторного часового ряду [7]:

$$Z = \{(t_i, V_i, S_i, P_i)\}_{i=1}^N \quad (1)$$

де t_i — часова мітка i -го вимірювання, V_i — рівень палива у баках (літри), S_i — швидкість тепловоза (км/год), P_i — потужність дизель-генераторної установки (кВт), N — загальна кількість точок вимірювання.

Для усунення високочастотного шуму та виявлення нефізичних стрибків сигналу обчислюється дискретна швидкість зміни рівня палива у баках — відношення приросту рівня палива до приросту часу між сусідніми відліками, л/хв:

$$\Delta V_i = \left| \frac{V_i - V_{i-1}}{t_i - t_{i-1}} \right| \quad (2)$$

Точки, у яких значення цієї швидкості, згладжене у вікні 20 відліків, за модулем перевищує робочий поріг 9 000 л/хв, вилучаються з ряду як інструментальні викиди й замінюються інтерпольованими значеннями: за такої швидкості весь паливний бак (номінальною ємністю 6 000 л) мав би спорожніти або наповнитися менш ніж за хвилину — значно швидше за реальний технологічний час заправки чи витрати, що вказує на апаратний артефакт, а не фізичну подію. Додатково, без зміни робочого порога, обраховується кількість

перевищень за нижчими діагностичними порогоми 500, 1 000 і 2 000 л/хв — щоб оцінити, скільки реальних подій жорсткіший поріг вирізав би як інструментальні викиди. Згладжування ряду виконується фільтром зі скінченною імпульсною характеристикою та гаусовим ваговим вікном:

$$V_i^* = \sum_{k=-m}^m w_k V_{i+k} \quad (3)$$

де w — вагові коефіцієнти гаусового вікна, нормовані так, що їх сума дорівнює одиниці; k — індекс відліку в межах вікна; m — половина його довжини. Довжина вікна завжди непарна, її мінімальне значення становить 5 відліків, а фактичне — задається адаптивно за типом експлуатаційного сегмента, середньою швидкістю руху та рівнем заповнення бака: на сегментах руху вікно найширше, на ділянках заправок — найвужче. Таке диференційоване згладжування, яке зберігає різкість фронту підйому рівня під час заправки, та критерії виявлення подій заправки описано нижче, у розділі «Критерії визначення меж експлуатаційних сегментів». На початку та в кінці ряду вікно згладжування виходить за межі наявних даних. Щоб уникнути спотворення на краях, ряд перед першим відліком і після останнього продовжується сталим значенням (відповідно першого й останнього відліку) на відрізок довжиною $2 \cdot \text{win}$ відліків, де win — довжина вікна згладжування, застосованого на цій ділянці.

Відновлення втрачених значень на ділянках мікророзривів, тривалістю до 10 хвилин, здійснюється методом лінійної інтерполяції; обмеженість такої схеми для тривалих розривів показано в [5]. Відновлене значення V_{int} у момент часу t на інтервалі між відомими точками (t_1, V_1) та (t_2, V_2) обчислюється за формулою:

$$V_{int}(t) = V_1 + \frac{(V_2 - V_1)}{(t_2 - t_1)} (t - t_1) \quad (4)$$

Інтерполяція забезпечує неперервність ряду рівня палива, що потрібно для наступного етапу — алгоритмічної сегментації [6]. Сегментація розділяє неперервний ряд на окремі експлуатаційні стани за потужністю дизель-генераторної установки: рух — потужність перевищує 0,1 кВт, стоянка — потужність не перевищує цього значення; окремо виділяється екіпірування. Енергія W , кВт·год, вироблена дизель-генераторною установкою на кожному сегменті, обчислюється шляхом чисельного інтегрування часового ряду потужності за методом трапецій: $W = \sum (P_k + P_{k+1}) \cdot \Delta t / 2$, де P_k — потужність на k -му відліку сегмента, кВт, а $\Delta t = 1/60$ год — крок дискретизації.

Критерії визначення меж експлуатаційних сегментів

Виявлення заправок виконується за додатним викидом згладженої швидкості зміни рівня палива, обчисленої за формулою (2), який перевищує локальну оцінку рівня завад, помножену на коефіцієнт 3. Ця оцінка формується як зважена сума трьох складників:

- Перший — модуль різниці значень швидкості руху між двома сусідніми відліками, пропорційний прискоренню, згладжений тим самим гаусовим вікном, що й ряд рівня палива. Цей доданок входить до суми порогу з додатним знаком, тому на ділянках розгону й гальмування, де різниця швидкостей більша, порогове значення пропорційно зростає.
- Другий — потужність дизель-генераторної установки, згладжена тим самим вікном і зважена власним масштабним коефіцієнтом; цей доданок також входить до суми з додатним знаком, тому на ділянках із вищою потужністю, тобто під навантаженням, порогове значення пропорційно зростає.
- Третій — квадратний корінь із ковзної дисперсії квадрата тієї самої згладженої швидкості зміни рівня палива, обчисленої у вікні 80 відліків; він зростає там, де ця швидкість нестабільна від відліку до відліку.

Кожен складник входить у суму з власним масштабним коефіцієнтом, підібраним емпірично, а сама сума обмежена знизу, щоб на «тихих» ділянках поріг не спадав до нуля. Завдяки цьому поріг є адаптивним: він змінюється разом із локальним рівнем шуму та динамікою руху, а не є сталою величиною. Подія визнається заправкою, якщо приріст рівня палива перевищує 300 л; суміжні події, розділені інтервалом до 10 хвилин, об'єднуються в одну.

Навколо кожної заправки встановлюється зона виключення тривалістю 5 хвилин до її початку та 25 хвилин після завершення: у цьому інтервалі рівень у баку осідає після наливу, тому зміни рівня палива в ньому не враховуються під час визначення меж сегментів. Мінімальна амплітуда зміни рівня палива, що враховується, становить 20 л.

Кожна виділена зміна рівня палива перевіряється на перевищення шумового фону: відношення її амплітуди до локального середньоквадратичного відхилення має бути не меншим за три, інакше зміна вважається шумом і межа сегмента не встановлюється. Додатково враховуються лінійність та однорідність зміни, частка часу руху об'єкта в її межах і повернення рівня палива до попереднього значення після завершення.

Статистичний аналіз суцільного часового ряду

Дослідження охоплює суцільну вибірку з шести маневрових тепловозів загальним обсягом 4 898 881 хвилинних вимірювань. Прийнята в розрахунках ємність паливного бака становить 6 000 л. Розраховані за стандартними означеннями [8] глобальні статистичні характеристики масиву наведено в табл. 1.

Коефіцієнт асиметрії ряду рівня палива лежить у межах від -0,510 до -0,247, що відповідає помірній лівій асиметрії розподілу. Зафіксовано 325 мікророзривів загальним обсягом 1 099 відліків — саме вони підлягають відновленню лінійною інтерполяцією. Разом з тим, наявність 53 макророзривів, максимальна зафіксована тривалість відсутності зв'язку в яких сягала 195 діб, обґрунтовує блокування наскрізної інтерполяції для тривалих періодів, оскільки за таких інтервалів відновлення проміжних значень не має фізичного підґрунтя.

Таблиця 1

Статистичні характеристики неперервного масиву телеметричних даних

Характеристика	Значення
Загальна кількість точок вимірювання	4 898 881
Базовий медіанний крок дискретизації	60 с
Кількість точок нижче робочого діапазону датчика («мертва зона»)	97 494 відліки
Кількість інтерпольованих мікророзривів тривалістю не більше 10 хв	325 випадків — 1 099 відліків
Кількість макророзривів без наскрізної інтерполяції	53 випадки
Коефіцієнт асиметрії ряду рівня палива	від -0,510 до -0,247
Експес ряду рівня палива	від 2,14 до 4,35

Структура пропусків у телеметричному масиві

Перед обґрунтуванням параметрів відновлення доцільно розділити причини появи пропусків: у телеметричному потоці вони мають різну природу й потребують різної обробки. У досліджуваному масиві виділено три такі джерела.

- Перше — відліки, що лежать нижче відомого порогового рівня заповнення бака, за який рівень палива фізично не опускається; вони формально присутні у файлі, проте не несуть коректної інформації про рівень палива. При подальшій обробці такі відліки замінюються інтерпольованими значеннями за тим самим принципом, що й інші інструментальні викиди, і в такому вигляді входять до робочого ряду.
- Друге — мікророзриви часової сітки, тобто короточасні втрати зв'язку тривалістю не більше 10 хвилин, за яких відліки відсутні взагалі; саме вони підлягають відновленню.
- Третє — макророзриви тривалістю понад 10 хвилин, за яких відліки також відсутні, але відновлення не виконується; їх походження розглянуто нижче.

Межа 10 хвилин задана як параметр алгоритму обробки й узгоджена з двома його власними параметрами. По-перше, за найвищої зафіксованої в парку середньої витрати на сегментах руху, яка становить 91 л/год, розрив тривалістю 10 хвилин відповідає зміні рівня приблизно на 15 л — це менше за мінімальну амплітуду зміни рівня палива, яку алгоритм узагалі враховує (20 л). Отже, наскрізна інтерполяція такого розриву не здатна ані створити, ані приховати межу сегмента за рахунок звичайної витрати. По-друге, різка подія всередині такого розриву не втрачається: перепад рівня між крайніми точками зберігається, а зміщується лише часова прив'язка події — не більш як на тривалість розриву, тобто в межах того самого десятихвилинного вікна об'єднання суміжних подій, яке алгоритм застосовує і без інтерполяції. Розриви понад цю межу мають неоднорідне походження: найкоротші з них, від 11 хвилин, — це триваліші втрати зв'язку, а найдовші, до 195 діб, відповідають паузам між сеансами вивантаження телеметрії та зняття локомотива з експлуатації; для всіх них наскрізна інтерполяція не застосовується. Перевірку точності відновлення на цій межі наведено нижче, в розділі «Валідація методу відновлення пропусків». Розподіл мікророзривів за тривалістю наведено в табл. 2.

Таблиця 2

Розподіл мікророзривів телеметричного ряду за тривалістю — парк із шести локомотивів

Тривалість розриву, хв	Кількість розривів	Частка від загальної кількості, %	Відновлено відліків
1	43	13,2	43
2	138	42,5	276
3	39	12,0	117
4	22	6,8	88
5	5	1,5	25
6	11	3,4	66
7	53	16,3	371
8	13	4,0	104
9	1	0,3	9
Разом	325	100,0	1 099

Розподіл є істотно нерівномірним. Найчастішими є двохвилинні розриви — 42,5 % усіх випадків; разом з одно- та трихвилинними вони становлять 67,7 % загальної кількості. Проте за обсягом відновлених відліків картина інша: на семихвилинні розриви, яких за числом лише 16,3 %, припадає 371 відлік, тобто понад третину всього відновленого обсягу. Найдовший зафіксований мікророзрив має тривалість 9 хвилин, тому для цього масиву прийнята межа не стала обмежувальною: жоден короточасний розрив не було відкинуто через

її перевищення. Сумарний обсяг втручання в дані на цьому етапі становить 1 099 відліків із 4 898 881, тобто 0,022 % масиву.

Розподіл пропусків за окремими об'єктами наведено в табл. 3.

Таблиця 3

Структура пропусків у телеметричних даних за об'єктами

Об'єкт	Мікророзрив	Відновлено відліків	Точок нижче робочого діапазону	Макророзрив	Максимальний розрив	Перевишень діагностичного порога 1 000 л/хв
1	20	61	43 962	15	122 год	4
2	0	0	348	1	1,4 год	0
3	50	152	7 212	9	195 діб	0
4	131	602	23 480	14	10,4 год	0
5	113	246	12 007	8	122 год	0
6	11	38	10 485	6	18,9 діб	0
Разом	325	1 099	97 494	53	195 діб	4

Наведені дані показують, що переважна частина неоднорідності масиву пов'язана не з перервами зв'язку, а з показаннями датчика поза його робочим діапазоном: таких показань 97 494, тоді як реально відсутніх у часовій сітці відліків — 1 099, тобто у 89 разів менше. Частка реально відсутніх відліків становить лише 0,022 % загального обсягу вибірки. Об'єкти суттєво різняться за якістю каналу зв'язку: в об'єкта 2 мікророзривів не зафіксовано взагалі, тоді як в об'єкта 4 їх 131. Максимальні зафіксовані макророзриви — від 1,4 години до 195 діб — відповідають плановим паузам в експлуатації та між сеансами вивантаження телеметрії. Відновлювати такі інтервали інтерполяцією неможливо в принципі: за час відсутності зв'язку локомотив міг бути запровлений, переміщений або виведений у ремонт, і жодна апроксимація за двома крайніми точками цих подій не відтворює.

Валідація методу відновлення пропусків

Обґрунтованість застосування лінійної інтерполяції та вибір граничної тривалості мікророзриву перевірено обчислювальним експериментом на реальних даних. На кожному об'єкті випадковим чином обиралися ділянка неперервного ряду заданої тривалості, проміжні відліки якої тимчасово вилучалися та відновлювалися лінійною інтерполяцією за формулою (4) між крайніми відомими точками. Функція інтерполяції отримувала лише два крайні відомі значення інтервалу; приховані проміжні значення їй не передавалися і використовувалися тільки для порівняння з відновленими. Для кожного об'єкта та кожної граничної тривалості виконувалося 1000 незалежних спроб; генератор випадкових чисел визначає, яку саме позицію в ряду буде обрано для приховування на кожній спробі, а фіксація його початкового стану означає, що при повторному запуску експерименту буде обрано ту саму послідовність позицій і, отже, отримано ті самі числові результати. До статистики зараховувалися лише ті спроби, для яких у вибраній позиції існував суцільний інтервал заданої тривалості без реальних пропусків; решта відкидалася, тому кількість зарахованих випробувань у табл. 4 становить від 997 до 1000. Експеримент виконувався після відсікання «мертвої зони» датчика, але до згладжування, тобто оцінює похибку саме інтерполяції, а не всього ланцюга обробки. Оскільки для випробувань придатні лише ділянки, вільні від реальних пропусків, одержані оцінки слід трактувати як нижню межу похибки.

Точність відновлення оцінювали двома показниками: середньою абсолютною похибкою MAE (mean absolute error), вираженою в літрах, та середньою абсолютною відсотковою похибкою MAPE (mean absolute percentage error), що нормує похибку на фактичний рівень палива у баку. Результати для граничної тривалості розриву 10 хвилин наведено в табл. 4.

Середня абсолютна похибка відновлення не перевищує 3,4 л для жодного з об'єктів парку, а найбільший її 95-й перцентиль — 10,3 л, що відповідає відносній похибці від 0,013 до 0,108 %.

Розкид похибки MAE між об'єктами (табл. 4) становить 8,5 рази і не є випадковим: він статистично пов'язаний із двома характеристиками самих об'єктів. Об'єкти з більшою часткою часу руху систематично показують більшу похибку відновлення — коефіцієнт кореляції між часткою часу руху та середньою абсолютною похибкою становить 0,95, а з відносною похибкою MAPE — 0,98. Аналогічно об'єкти з вищим середньоквадратичним відхиленням шуму на стоянці систематично показують більшу похибку відновлення — коефіцієнт кореляції становить 0,90. Найменшу похибку показав об'єкт 6, який перебуває в русі лише 0,66 % часу, найбільшу — об'єкт 2 із часткою руху 22,6 %. Це відповідає фізиці процесу: на стоянці рівень палива змінюється повільно й майже лінійно, тому пряма між крайніми відомими точками відтворює його точно, тоді як у русі додаються витрата під навантаженням і перехідні процеси, які порушують лінійність.

Показовим є відношення похибки відновлення (MAE, табл. 4) до середньоквадратичного відхилення шуму на стоянці для кожного об'єкта; це відхилення, обчислене в рамках тої самої обробки даних, становить за об'єктами відповідно 12,6; 20,8; 27,1; 12,2; 20,3 та 4,8 л. За об'єктами таке відношення становить 0,071–

0,163, тобто розкид скорочується з 8,5 до 2,3 рази. Інакше кажучи, інтерполяція поводить на всіх об'єктах однаково і відновлює рівень палива з точністю близько десятої частини локального шуму, а різницю між об'єктами визначає не метод, а якість вимірювального каналу та режим експлуатації. Наведені кореляції обчислено за шістьма об'єктами, тому їх слід трактувати як вказівку на зв'язок, а не як його доведення. Зіставлення з мінімальним обсягом заправки, що підлягає виявленню (300 л), показує, що похибка інтерполяції на два порядки менша за поріг детекції події. Отже, відновлення мікророзривів не здатне ані породити хибне спрацювання алгоритму детекції заправок, ані замаскувати реальну подію.

Таблиця 4

Похибка відновлення пропусків лінійною інтерполяцією за граничної тривалості розриву 10 хвилин

Об'єкт	Кількість випробувань	MAPE		MAE	
		середня, %	медіанна, %	середня, л	95-й процентиль, л
1	997	0,030	0,007	1,35	5,11
2	1000	0,108	0,039	3,39	6,41
3	1000	0,079	0,024	2,77	7,36
4	1000	0,024	0,007	0,87	3,09
5	999	0,069	0,016	2,62	10,25
6	998	0,013	0,000	0,40	0,80

Для перевірки того, наскільки критичним є вибір саме 10-хвилинної межі, той самий експеримент із прихованням і відновленням значень повторено для граничних тривалостей 5, 15, 20 і 30 хвилин. Крім лінійної інтерполяції, на тих самих прихованих ділянках перевірено ще дві прості схеми відновлення: заміну прихованих значень найближчим відомим (метод «найближчий сусід») та утримання останнього відомого значення до кінця розриву (метод «утримання останнього»). Результати наведено в табл. 5.

Таблиця 5

Залежність похибки відновлення від граничної тривалості розриву та методу інтерполяції (середнє за парком)

Гранична тривалість розриву, хв	Лінійна		Найближчий сусід		Утримання останнього	
	MAE, л	MAPE, %	MAE, л	MAPE, %	MAE, л	MAPE, %
5	1,37	0,039	1,60	0,045	2,11	0,059
10	1,90	0,054	2,37	0,067	3,59	0,095
15	2,54	0,067	3,24	0,085	5,35	0,131
20	3,25	0,100	3,74	0,109	5,91	0,152
30	3,34	0,090	3,88	0,106	5,93	0,157

За середньою абсолютною похибкою (MAE) зростання для всіх трьох схем монотонне, проте для лінійної інтерполяції навіть на 30-хвилинних ділянках вона залишається в межах 3,3 л. Відносна похибка (MAPE) лінійної схеми зростає монотонно від 5 до 20 хвилин (0,039–0,100 %), але на 30-хвилинному порозі знижується до 0,090 %; це не суперечить висновку про консервативність межі 10 хвилин, проте ця відмінність (0,100 % проти 0,090 %) не перевищує похибки середнього за парком (відповідно 0,034 та 0,020 відсоткових пункти для 6 локомотивів) і тому не є статистично значущою: перевірка на межі ± 1 похибки середнього підтверджує, що на всіх п'яти порогах і для всіх трьох перевірених методів немонотонності, яка виходила б за межі похибки вибірки, не виявлено. Прийнята межа 10 хвилин є, таким чином, консервативною і забезпечує майже дворазовий запас за точністю відносно верхньої перевіреної тривалості. Блокування наскрізної інтерполяції для тривалих розривів зумовлене не похибкою самої апроксимації, а принциповою неможливістю відновити за двома крайніми точками експлуатаційні події, що відбулися всередині багатогодинної «сліпої зони».

Лінійна інтерполяція стійко переважає обидві альтернативні схеми за абсолютною похибкою (MAE): заміна найближчим відомим значенням дає похибку в середньому на 15–28 % більшу, а утримання останнього відомого значення — у 1,5–2,1 рази. За відносною похибкою (MAPE, табл. 5) така сама відмінність зберігається на всіх п'яти перевірених порогах: найближчий сусід дає MAPE на 9–28 % більший, а схема утримання останнього значення — у 1,5–2 рази більший за лінійну, тобто та сама залежність, що й за MAE. Кубічний сплайн у цьому експерименті чисельно збігається з лінійною схемою: він також будується лише за двома крайніми відомими точками розриву — тим самим локальним підходом, який реалізує лінійна інтерполяція, — а за двома точками кубічний сплайн вироджується в пряму. Це узгоджено з методом, застосованим у виробничому конвеєрі (заповнення пропусків за двома найближчими відомими сусідніми точками): порівняння методу, що використовує ширше вікно опорних точок, перевіряло би вже іншу схему відновлення і виходить за межі цього дослідження.

Одержані залежності дають орієнтир для вибору межі на іншому масиві. Допустимість інтерполяції визначається не абсолютною похибкою відновлення, а її співвідношенням з порогом детекції подій.

Відповідно граничну тривалість розриву слід обирати так, щоб середня абсолютна похибка не перевищувала десятої частини порога детекції. Для розглянутого парку, де поріг детекції заправки становить 300 л, а середня абсолютна похибка на 30-хвилинних розривах — 3,3 л, ця умова виконується з великим запасом навіть на верхній перевірній межі. Отже, обмеження 10 хвилин продиктоване не точністю апроксимації, а обережністю щодо експлуатаційних подій усередині розриву: чим довший інтервал, тим вища ймовірність, що в ньому відбулася заправка, і тим більший ризик, що лінійна апроксимація згладить її у монотонний тренд.

Статистика експлуатаційних сегментів та інтерполяції

Після відновлення цілісності ряду за допомогою інтерполяції масив даних було розділено на 67 855 окремих експлуатаційних сегментів (табл. 6). Аналіз сегментованих даних дозволяє детально дослідити структуру енергоспоживання. Статистика підтверджує, що маневрові тепловози більшу частину часу проводять у стані простою — від 57,7 до 98,8 % загального часу за окремими об'єктами. Проте аналіз сегментів стоянки показує, що дизель-генераторна установка продовжує працювати, генеруючи середню потужність на рівні 4,9–11,0 кВт для забезпечення власних потреб локомотива.

Таблиця 6

Статистичний розподіл виділених експлуатаційних сегментів по парку

Об'єкт	Кількість сегментів загалом	Кількість ідентифікованих заправок
1	7858	23
2	3821	8
3	20312	118
4	13365	46
5	22163	82
6	336	3
Разом	67 855	280

Розподіл виділених сегментів за експлуатаційними станами наведено в табл. 7.

Таблиця 7

Розподіл сегментів за типами експлуатаційних станів по парку

Тип сегмента	Кількість	Сумарна тривалість, хв	Частка часу, %	Витрата палива, л	Середня потужність, кВт
Рух	41 733	781 620	10,35	767 745	142,8
Стоянка	25 842	6 686 585	88,56	188 929	8,6
Екіпірування (заправка)	280	81 990	1,09	—	2,4
Разом	67 855	7 550 195	100,00	956 674	—

Структура експлуатаційного часу різко асиметрична: на стоянку припадає 88,6 % загального фонду часу при 10,4 % на рух, тоді як за обсягом спожитого палива співвідношення протилежне — 19,8 % проти 80,2 %. Середня потужність дизель-генераторної установки на стоянці становить 8,6 кВт проти 142,8 кВт у русі, що відповідає роботі установки на власні потреби локомотива. Сегменти екіпірування за визначенням не дають витрати палива й виділяються окремо, оскільки їх включення до балансу спотворило б оцінку витрати.

Аналіз кореляції між обсягом витраченого палива та виробленою дизель-генераторною установкою енергією на сегментах руху демонструє слабкий позитивний лінійний зв'язок на рівні 0,05–0,16. Такий рівень зв'язку означає, що поточна потужність установки пояснює лише незначну частку варіації витрати палива на окремому сегменті. Це робить ненадійним оцінювання витрати за поточковими різницями рівня палива й обґрунтовує перехід до інтегрального оцінювання на рівні сегментів. Установлення конкретних причин слабого зв'язку виходить за межі цієї роботи.

Вплив відновлення пропусків на результат сегментації

Щоб кількісно оцінити внесок відновлення точок «мертвої зони» у кінцевий результат, увесь ланцюг оброблення — сегментацію та детекцію заправок — виконано двічі: з лінійною інтерполяцією 97 494 точок, що лежать нижче робочого діапазону, і без неї, коли ці точки залишено як пропущені значення. Мікророзриви часової сітки — 325 випадків і 1 099 відліків в обох варіантах цього порівняння не заповнюються: вони є відсутніми рядками, а не значеннями в масиві, тому окремо в цьому експерименті не тестуються. Порівняння наведено в табл. 8.

Результат такий: заповнення 97 494 точок «мертвої зони», тобто 2,0 % масиву, змінило кількість заправок лише на одну — з 280 до 281, і лише для об'єкта 3, де без заповнення виявлено 118 заправок, а із заповненням — 119; для решти п'яти об'єктів кількість заправок не змінилася. Різниця в кількості виділених сегментів становить 10 одиниць, тобто 0,015 %. Незначна розбіжність загальної кількості сегментів порівняно з табл. 6 зумовлена повторним проходом детекції в межах контрольного експерименту і на висновок не впливає.

Таблиця 8

Порівняння результатів сегментації та детекції заправок без відновлення пропусків і з відновленням

Об'єкт	Різниця в кількості виділених сегментів	Заправок (без / з)
1	+4	23 / 23
2	0	8 / 8
3	+4	118 / 119
4	-2	46 / 46
5	+4	82 / 82
6	0	3 / 3
Разом	+10	280 / 281

Отриманий результат означає, що заповнення точок «мертвої зони» лінійною інтерполяцією не спричиняє суттєвих змін результату сегментації та детекції заправок навіть тоді, коли обсяг заповнюваних точок сягає 2 % масиву. Цей експеримент перевіряє саме заповнення точок нижче робочого діапазону; вплив відновлення саме мікророзривів часової сітки перевірено окремим тестом: часову сітку реконструйовано, вставивши в кожен мікророзрив відсутні відліки з лінійно інтерпольованими між сусідніми точками рівнем палива, потужністю та швидкістю — 325 випадків, 1 099 вставлених відліків, точно відповідає раніше діагностованій кількості пропусків часової сітки — і весь ланцюг оброблення виконано повторно на цій реконструйованій осі. Для об'єкта 2 мікророзривів, що відповідають критерію, не виявлено. Для решти п'яти об'єктів кількість виділених сегментів змінилася на 0, +2, +33, +12 і +4 одиниці відповідно, разом +51 із 64 217 сегментів усіх типів, виділених у цьому прогоні за цими п'ятьма об'єктами, тобто 0,079 %; кількість заправок — лише для об'єкта 6, з 3 до 4. Це підтверджує висновок попереднього тесту: відновлення мікророзривів, як і заповнення «мертвої зони», не спричиняє суттєвих змін результату сегментації. Водночас цей тест — гіпотетичний: у виробничому конвеєрі мікророзриви часової сітки фактично не відновлюються, бо функція заповнення пропусків діє лише на наявні значення масиву й не бачить відсутніх рядків; тому наведена в роботі кількість заправок 280 і 281 відповідає поточній реалізації без реконструкції часової сітки.

Оцінка якості попередньої обробки сигналу

Ефективність попередньої обробки оцінено за зниженням рівня шуму на сегментах стоянки, де рівень палива фізично сталий, а отже будь-яка варіація показань є шумом: для кожного об'єкта обчислено середньоквадратичне відхилення показань на таких сегментах до та після адаптивного згладжування, а їх відношення виражено в децибелах. Другим показником є коефіцієнт детермінації балансової моделі витрати палива. Ця модель описує витрату як суму двох складників: питомої витрати, пропорційної потужності дизель-генераторної установки на активних хвилинах, і майже сталої витрати на холостому ході. Коефіцієнт детермінації обчислено за погодинними сумами фактичної та модельної витрати на згладженому ряді, з якого вилучено вікна заправок, і лише для тих годин, де щонайменше 75 % хвилин мають придатні відліки (табл. 9).

Таблиця 9

Показники якості попереднього оброблення телеметричного сигналу

Об'єкт	Зниження рівня шуму на стоянці, дБ	Коефіцієнт детермінації моделі витрати
1	0,29	0,194
2	6,10	0,521
3	3,17	0,433
4	0,57	0,319
5	1,20	0,357
6	-0,01	0,032
Середнє за парком	1,89	0,309

Зниження рівня шуму на сегментах стоянки становить від нуля до 6,1 дБ, у середньому 1,9 дБ за парком, що відповідає зменшенню амплітуди шумової складової приблизно на 20 %. Ефект істотно різниться між об'єктами: найбільший він для об'єктів 2 і 3 — 6,1 та 3,2 дБ відповідно, тоді як для об'єкта 6 згладжування рівня шуму практично не змінює. Це узгоджується з вихідним станом вимірювального каналу: об'єкт 6 має найнижче середньоквадратичне відхилення шуму на стоянці — 4,8 л, тому запасу для подавлення в нього майже немає, тоді як об'єкти з "найгучнішим" каналом, де це відхилення становить 20,8 та 27,1 л, отримують найбільший вигаш.

Коефіцієнт детермінації балансової моделі витрати палива залишається помірним — від 0,03 до 0,52 за середнього значення 0,31. Це узгоджується зі слабким лінійним зв'язком між обсягом витраченого палива та виробленою енергією, зафіксованим вище, і свідчить, що поточна потужність дизель-генераторної установки не є достатнім предиктором миттєвої витрати палива. Практичний наслідок полягає в тому, що оцінювати витрату палива слід інтегрально, на рівні експлуатаційних сегментів, а не за поточковими різницями рівня палива.

Обмеження запропонованого підходу

Отримані результати мають низку обмежень, які слід урахувати під час їх інтерпретації та перенесення на інші масиви. По-перше, оцінка похибки відновлення, наведена в табл. 4 і 5, є нижньою межею:

для випробувань відбиралися виключно ділянки ряду, вільні від реальних пропусків, тобто заздалегідь якісніші за середню ділянку масиву. По-друге, порівняння схем інтерполяції виконано на конфігурації з двома опорними точками — крайніми відомими значеннями розриву. За такої постановки кубічний сплайн вироджується в пряму і чисельно збігається з лінійною схемою, тому висновок про перевагу лінійної інтерполяції є коректним лише щодо схем заміни найближчим значенням та утримання останнього відомого значення.

По-третє, робочий критерій вилучення нефізичних стрибків за дискретною швидкістю зміни рівня палива (поріг 9 000 л/хв) на цій вибірці не спрацював ані разу, а за нижчим діагностичним порогом 1 000 л/хв зафіксовано лише 4 перевищення, і всі — на одному об'єкті, тоді як 97 494 відліки замінені інтерпольованими значеннями за ознакою виходу за відомий поріг заповнення бака. Це означає, що для цієї вибірки основним джерелом заміщених значень є не імпульсний шум, а вихід показань за заданий поріг робочого діапазону, а поріг критерію імпульсних викидів потребує окремого калібрування під конкретний тип обладнання.

По-четверте, макророзриви тривалістю до 195 діб не піддаються відновленню жодним із розглянутих методів, а їх наявність унеможливляє побудову суцільного паливного балансу за весь період спостереження. По-п'яте, оцінки якості попередньої обробки отримано на вибірці з шести об'єктів; для узагальнення на інші серії рухомого складу потрібна ширша вибірка. При цьому з шести об'єктів лише п'ять мають високу оцінку достовірності даних: об'єкт 6 позначено як такий, що вимагає перевірки датчика, і саме він дає найнижчий коефіцієнт детермінації моделі витрати — 0,032 — і нульовий вигравш від згладжування, тому його найменші показники похибки відновлення не слід трактувати як найліпшу якість вимірювального каналу.

По-шосте, тест впливу відновлення саме мікророзривів часової сітки є гіпотетичним: у виробничому конвеєрі таке відновлення фактично не виконується — його реалізовано окремо, лише для цього контрольного тесту, тому наведена в роботі кількість сегментів і заправок відповідає поточній реалізації виробничого конвеєра, без такої реконструкції. По-сьоме, сумарна тривалість сегментів у табл. 7, а це 7 550 195 хв, перевищує загальну кількість вимірювань 4 898 881 через виявлену ваду конвеєра складання вхідних файлів: контрольна діагностика показала, що синтетична тривалість перевищує реальний календарний діапазон на 13–80 % залежно від об'єкта, ймовірно через файли з перекривними або непорядкованими датами, які складаються як послідовні. Кількість сегментів 67 855 і кількість заправок 280 і 281 від цієї вади не залежать і залишається надійною; частку часу за типами сегментів (табл. 7) слід трактувати з обережністю до усунення вади в конвеєрі. Нарешті, окреме обмеження стосується меж сегментів екіпірування: сумарна тривалість 280 сегментів заправки становить 81 990 хвилин, тобто в середньому близько 4,9 години на подію, що істотно перевищує технологічний час наливу. Причина такої тривалості не встановлена й потребує окремої перевірки: вона може бути пов'язана як із правилом об'єднання близько розташованих подій, так і з фіксацією межі сегмента після завершення перехідного процесу осідання рівня палива. На оцінювання витрати це не впливає, оскільки вікно заправки виключається з аналізу, проте для оцінювання фактичної тривалості екіпірування такі межі потребують уточнення.

ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМІ

Проведений статистичний аналіз масиву обсягом 4,89 млн телеметричних вимірювань дозволив визначити статистичні характеристики рядів рівня палива маневрових тепловозів. Застосування згладжування гаусовим ваговим вікном та лінійної інтерполяції дозволило відновити цілісність часового ряду для 325 мікророзривів зв'язку загальним обсягом 1 099 відліків, замінити інтерпольованими значеннями 97 494 відліки, що лежали нижче робочого діапазону датчиків рівня палива, та зафіксувати 4 перевищення діагностичного порога швидкості зміни рівня палива 1 000 л/хв, усі на одному об'єкті. Подальша алгоритмічна сегментація очищеного ряду дозволила виділити майже 68 тисяч експлуатаційних станів. Контрольні експерименти показали, що на досліджуваному масиві внесок відновлення пропусків у результат сегментації є практично нульовим: заповнення 2,0 % масиву нижче робочого діапазону датчика змінило кількість сегментів на 0,015 % і заправок — лише на одну з 280; гіпотетичне відновлення мікророзривів, тобто 0,022 % масиву, змінило кількість сегментів на 0,079 % і заправок — теж лише на одну. Роль відновлення пропусків у ланцюзі «фільтрація — інтерполяція — сегментація» полягає в забезпеченні неперервності ряду, потрібної для обчислення швидкості зміни рівня палива та визначення меж сегментів. Сукупно це дає набір показників якості попередньої обробки сигналу, який можна переносити на інші масиви: точність відновлення на прийнятій межі — від 0,4 до 3,4 л за MAE і від 0,013 до 0,108 % за MAPE; відношення цієї похибки до локального шуму, що становить від 0,071 до 0,163, та до порога детекції заправки, а також частку відліків, що зазнали втручання: 2,0 % за ознакою виходу за робочий діапазон та 0,022 % за ознакою мікророзривів часової сітки.

Перспективи подальших досліджень у даному напрямі полягають у розширенні набору методів відновлення пропусків за рахунок нелінійних та модельно-орієнтованих схем імпутизації для макророзривів, в уточненні критеріїв визначення меж експлуатаційних сегментів за сукупністю ознак (потужність, швидкість,

тривалість сегмента), а також у перевірці запропонованого комбінованого підходу на розширеній вибірці рухомого складу інших серій та в режимі обробки телеметрії у реальному часі.

Література

1. Delussu, F., Imran, F., Mattia, C., Meo, R. Fuel prediction and reduction in public transportation by sensor monitoring and Bayesian networks. *Sensors*. 2021. Vol. 21, no. 14. P. 4733. <https://doi.org/10.3390/s21144733>
2. Navixy. Types of fuel level sensors: resistive, capacitive, ultrasonic methods. URL: <https://navixy.com/docs/expert-center/vehicle-telematics-technology/fuel-management/installation-and-initial-configuration-of-fuel-control-devices/fuel-level-sensors/types-of-fuel-level-sensors>
3. Terzic, E., Nagarajah, C. R., Alamgir, M. Capacitive sensor-based fluid level measurement in a dynamic environment using neural network. *Engineering Applications of Artificial Intelligence*. 2010. Vol. 23, no. 4. P. 614–619. <https://doi.org/10.1016/j.engappai.2009.09.014>
4. НВО «Дніпротехтранс». Вимірювання витрати палива датчиками рівня палива у баці. URL: <http://dneprroteh.com/?p=9074>
5. Liu, Y., Dillon, T., Yu, W., Rahayu, W., Mostafa, F. Missing value imputation for industrial IoT sensor data with large gaps. *IEEE Internet of Things Journal*. 2020. Vol. 7, no. 8. P. 6855–6867. <https://doi.org/10.1109/JIOT.2020.2970467>
6. Івашев Д., Герасимов В. Алгоритм сегментації даних вимірювань рівня палива на транспорті. Вимірювальна та обчислювальна техніка в технологічних процесах. 2025. Т. 82, № 2. С. 419–425. <https://doi.org/10.31891/2219-9365-2025-82-60>
7. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., Ljung, G. M. *Time Series Analysis: Forecasting and Control*. 5th ed. Hoboken, NJ: John Wiley & Sons, 2015. 712 p.
8. Montgomery, D. C., Jennings, C. L., Kulahci, M. *Introduction to Time Series Analysis and Forecasting*. 2nd ed. Hoboken, NJ: Wiley, 2015. 672 p.

References

1. Delussu F., Imran F., Mattia C., Meo R. Fuel prediction and reduction in public transportation by sensor monitoring and Bayesian networks. *Sensors*. 2021. Vol. 21, no. 14. P. 4733. <https://doi.org/10.3390/s21144733>
2. Navixy. Types of fuel level sensors: resistive, capacitive, ultrasonic methods. URL: <https://navixy.com/docs/expert-center/vehicle-telematics-technology/fuel-management/installation-and-initial-configuration-of-fuel-control-devices/fuel-level-sensors/types-of-fuel-level-sensors>
3. Terzic E., Nagarajah C. R., Alamgir M. Capacitive sensor-based fluid level measurement in a dynamic environment using neural network. *Engineering Applications of Artificial Intelligence*. 2010. Vol. 23, no. 4. P. 614–619. <https://doi.org/10.1016/j.engappai.2009.09.014>
4. NVO "Dniprotekhtrans". Vymiruvannya vytraty palyva datchykamy rivnia palyva u bantsi [Measuring fuel consumption using fuel level sensors in the tank]. URL: <http://dneprroteh.com/?p=9074> (in Ukrainian)
5. Liu Y., Dillon T., Yu W., Rahayu W., Mostafa F. Missing value imputation for industrial IoT sensor data with large gaps. *IEEE Internet of Things Journal*. 2020. Vol. 7, no. 8. P. 6855–6867. <https://doi.org/10.1109/JIOT.2020.2970467>
6. Ivashchev D., Herasymov V. 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*. 2025. Vol. 82, no. 2. P. 419–425. <https://doi.org/10.31891/2219-9365-2025-82-60> (in Ukrainian)
7. Box G. E. P., Jenkins G. M., Reinsel G. C., Ljung G. M. *Time Series Analysis: Forecasting and Control*. 5th ed. Hoboken, NJ: John Wiley & Sons, 2015. 712 p.
8. Montgomery D. C., Jennings C. L., Kulahci M. *Introduction to Time Series Analysis and Forecasting*. 2nd ed. Hoboken, NJ: Wiley, 2015. 672 p.