

УДК 004.62:504.3.054

МЕТОДИ ТА ПРОГРАМНІ ЗАСОБИ ПЕРВИННОЇ ОБРОБКИ ДАНИХ У ЗАДАЧАХ МОНІТОРИНГУ ЗАБРУДНЕННЯ АТМОСФЕРНОГО ПОВІТРЯ В ОКОЛІ ОБ'ЄКТІВ ЕНЕРГЕТИКИ

С. І. Ковтун¹, Ю. В. Куц², Б. Б. Млинко³, М. Є. Фриз⁴

^{1,2,4}Інститут загальної енергетики НАН України, вул. Антоновича, 172, м. Київ, 03150, Україна

²НТУУ "Київський політехнічний інститут ім. Ігоря Сікорського", Берестейський пр., 37, м. Київ, 03056, Україна

^{3,4}Тернопільський національний технічний університет імені Івана Пулюя, вул. Руська, 56, м. Тернопіль, 46001, Україна, тел.: +38(0352)519714, e-mail: ⁴mykh.fryz@gmail.com

Обґрунтовано вибір методів первинної обробки даних сенсорів в системі моніторингу якості атмосферного повітря в околі об'єктів енергетики. Запропоновано сучасні програмні інструменти, які дозволяють здійснювати таку обробку.

Ключові слова: енергоефективність, енергетична інформатика, моніторинг, забруднення повітря, аналіз даних, часові ряди.

METHODS AND SOFTWARE FOR PRIMARY DATA PROCESSING IN THE TASK OF AIR POLLUTION MONITORING IN THE VICINITY OF ENERGY FACILITIES

S. Kovtun¹, Y. Kuts², B. Mlynko³, M. Fryz⁴

^{1,2,4}General Energy Institute of National Academy of Sciences of Ukraine, 172 Antonovycha str., Kyiv, Ukraine, 03150

²National Techn. University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37 Pr. Beresteiskyi, Kyiv, Ukraine, 03056

^{3,4}Ternopil Ivan Puluj National Technical University, 56 Ruska str., Ternopil, Ukraine, 46001
phone: +38(0352)519714, e-mail: ⁴mykh.fryz@gmail.com

The choice of methods for primary processing of sensor data in the system for monitoring the quality of atmospheric air in the vicinity of energy facilities is justified. Modern software tools that allow such processing are proposed.

Key words: energy efficiency, energy informatics, monitoring, air pollution, data analysis, time series.

ORCID: ¹0000-0002-8493-9474; ²0000-0002-8493-9474; ³0000-0003-0780-5365; ⁴0000-0002-8720-6479.

Моніторинг забруднення атмосферного повітря в околі об'єктів енергетики, поряд із вирішенням задач охорони довкілля, є також одним із факторів підвищення енергоефективності, адже надає об'єктивні дані про наслідки роботи обладнання та дозволяє швидко виявляти режими, що супроводжуються перевитратами палива й зайвим власним споживанням. Моніторинг є джерелом великих масивів вимірювань, що опрацьовуються інформаційними системами, а методи аналізу даних використовуються для прогнозування викидів, виявлення аномалій, оптимізації керування в реальному часі та підтримки рішень, що одночасно підвищують енергоефективність і зменшують вплив на довкілля. Таким чином моніторинг забруднення атмосферного повітря в околі об'єктів енергетики тісно пов'язаний із задачами енергетичної інформатики.

Сенсорні системи моніторингу якості атмосферного повітря формують потоки даних, що мають типові особливості: високу частоту вимірювань, нерівномірність у часі, пропуски, дублікати, а також вплив шуму та збоїв живлення чи зв'язку. Первинні дані часто зберігаються у форматі CSV як найпростіший і найпоширеніший обмінний формат для інструментальних вимірювань. Водночас CSV не гарантує коректності типів, часових зон, одиниць вимірювання чи синхронізації. Тому первинна обробка є критичним етапом, що забезпечує придатність даних для подальшого аналізу, порівняння з нормативами, побудови трендів і виявлення аномалій.

Для вирішення завдань імпорту, очищення та аналізу сенсорних даних стандартом де-факто є мова програмування Python та її спеціалізована екосистема бібліотек. Вибір конкретних інструментів базується на їхній ефективності при роботі з часовими рядами. Методи та програмні засоби моніторингу якості атмосферного повітря в околі об'єктів енергетики повинні бути інтегровані в загальну концепцію дослідження функціонування об'єктів енергетики [1, 2]. На нашу думку доцільно використовувати такі методи програмні засоби первинної обробки.

Бібліотека pandas – це основний інструмент для роботи з табличними даними. Вона ідеально підходить для обробки файлів формату CSV завдяки вбудованим оптимізованим функціям читання. У контексті моніторингу якості

атмосферного повітря головною перевагою pandas є потужний функціонал для роботи з часовими рядами. Бібліотека matplotlib відповідає за створення статичних, анімованих та інтерактивних візуалізацій. Бібліотека stasmmodels – це інструмент, необхідний для більш глибокого статистичного аналізу на етапі оцінки якості даних. Він дозволяє розкласти часовий ряд на компоненти (тренд, сезонна компонента, залишок), що допомагає відрізнити реальні добові коливання рівня забруднення від апаратного шуму сенсора чи аномальних викидів. Крім того є можливість використовувати моделі з циклостаціонарними властивостями, які характерні як для процесів енергоспоживання так і для динаміки забруднень повітря [3, 4].

Попередня обробка даних відбувається у такій послідовності. Перший крок - коректне завантаження CSV-даних із явним визначенням стовпців часу, числових полів концентрацій (наприклад, PM2.5, PM10, NO2), одиниць вимірювання та розділювачів. Типові проблеми: різні формати часу, некоректні рядки (наприклад, "nan", "-", "error"), дублікати часових міток. На цьому ж етапі вводяться базові правила контролю якості: допустимі діапазони значень, маркування підозрілих вимірювань. Для сенсорів якості атмосферного повітря важливо відділити явно некоректні значення (негативні концентрації, технічні коди помилки) від потенційно реальних піків забруднення.

Сукупність методів попереднього опрацювання даних вказаними вище програмними засобами включає, перш за все, агрегацію, інтерполяцію, редискретизацію та візуалізацію реалізацій досліджуваних часових рядів.

Сенсори моніторингу зазвичай фіксують дані з високою частотою. Для аналізу часто необхідно перейти до нижчої роздільної здатності: погодинних, добових або тижневих значень. Агрегація вирішує це завдання шляхом обчислення статистичних характеристик у межах заданих часових вікон. Пропуски у даних сенсорів є неминучими: вони виникають через тимчасові збої обладнання, втрату пакетів при передачі даних або планові технічні обслуговування. Для коротких пропусків (до кількох хвилин) добре підходить лінійна інтерполяція. Для сигналів із вираженою нелінійною динамікою, наприклад із добовими піками концентрацій, ефективнішою є поліноміальна або сплайн-інтерполяція. Редискретизація - це зміна часового кроку вихідного ряду. Як правило, для даних задач використовується downsampling (коли частота зменшується, наприклад, з хвилинних до годинних даних).

Нарешті, візуалізація є невід'ємною складовою первинної обробки даних, оскільки дозволяє швидко виявити аномалії, тренди, пропуски та ефект від застосованих методів обробки. За допомогою matplotlib будуються лінійні графіки часових рядів концентрацій для кожного забруднювача, що дає змогу візуально порівняти сирі та оброблені дані. Гістограми, а також ядерні оцінки щільності розподілу дозволяють оцінити статистичні характеристики вимірювань і виявити мультимодальні розподіли, що можуть свідчити про різні джерела забруднення.

Таким чином, дані з сенсорів якості повітря, отримані у форматі CSV, вимагають обов'язкової підготовки перед їх використанням у системах моніторингу. Використання бібліотеки pandas дозволяє ефективно вирішити проблему нерівномірності надходження даних через процеси редискретизації та агрегації. Методи інтерполяції забезпечують неперервність часового ряду, усуваючи пропуски від тимчасових втрат зв'язку. Водночас візуалізація за допомогою matplotlib гарантує прозорість процесу очищення та допомагає виявити апаратні помилки. Запропонований програмно-методологічний стек є оптимальним, надійним та достатнім для побудови якісних процедур опрацювання даних, отриманих під час моніторингу забруднень повітря в околі об'єктів енергетики.

ПОСИЛАННЯ

1. В. П. Бабак, Ю. В. Куц, М. В. Мислович, М. Є. Фриз, and Л. М. Щербак, Об'єктно-орієнтована ідентифікація стохастичних шумових сигналів. Київ: Наукова думка, 2024. doi: <https://doi.org/10.15407/978-966-00-1883-9>.
2. V. Babak, A. Zaporozhets, S. Kovtun, Y. Kuts, M. Fryz, and L. Scherbak, "The Concept of Research of the Electric Power Facilities Functioning," in *Systems, Decision and Control in Energy VI. Studies in Systems, Decision and Control*, V. Babak and A. Zaporozhets, Eds., Cham: Springer Nature Switzerland, 2024, pp. 3–33. doi: 10.1007/978-3-031-68372-5_1.
3. L. M. Scherbak, M. Y. Fryz, and V. A. Hotovych, "Electricity consumption simulation using random coefficient periodic autoregressive model," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1254, no. 1, p. 12027, 2023, doi: 10.1088/1755-1315/1254/1/012027.
4. Mykhailo Fryz and Leonid Scherbak, "Properties of discrete-time conditional linear cyclostationary random processes in the problems of energy informatics," *Syst. Res. Energy*, no. 1 (72), pp. 72–79, Apr. 2023, doi: 10.15407/srenergy2023.01.072.

УДК 697.329

ДЕЯКІ СИСТЕМИ КОНТРОЛЮ ТА ДІАГНОСТУВАННЯ НАКОПИЧУВАЧІВ ЕЛЕКТРОЕНЕРГІЇ

В. М. Зварич^{1,2}, Ю. І. Гижко¹, Л. Б. Остапчук¹, Г. А. Кузік¹