



БАГАТОКРИТЕРІАЛЬНИЙ ВИБІР МЕТОДУ ОХОЛОДЖЕННЯ ШІ-ДАТА-ЦЕНТРУ З ІНТЕГРАЦІЄЮ ВІДНОВЛЮВАНИХ ДЖЕРЕЛ ТА SMART-GRID

MULTI-CRITERIA SELECTION OF AI DATA CENTER COOLING METHOD WITH RENEWABLE ENERGY AND SMART-GRID INTEGRATION

Артеменко С.В.¹

Artemenko S. V.¹

¹ Одеський національний технологічний університет, вул. Канатна, 112, м. Одеса, 65039, Україна

¹ Odesa National University of Technology, 112 Kanatna St., Odesa, 65039, Ukraine

ORCID: <https://orcid.org/0000-0002-1398-1472>

E-mail: sergey.artemenko@cloud.ontu.edu.ua

Copyright © 2026 by author and the journal “Automation of technological and business – processes”.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0>



DOI: [10.15673/atbp.v18i1.3501](https://doi.org/10.15673/atbp.v18i1.3501)

Анотація. Запропоновано методологію вибору оптимального методу охолодження дата-центру для штучного інтелекту (ШІ-дата-центру) з урахуванням інтеграції відновлюваних джерел енергії та smart-grid технологій через систему теплової акумуляції. Розроблена тришарова процедура прийняття рішень поєднує фільтр термодинамічної здійсненності, апостеріорний Парето-аналіз та нечітку багатокритеріальну згортку в архітектурі віртуальної енергетичної системи. Серед чотирьох розглянутих критеріїв особливе місце займає індекс smart-grid інтеграції, формалізований як приведена п'ятирічна вартість двох smart-grid сервісів: RTP-арбітражу через систему акумуляцію теплової енергії та утилізації скидної теплоти в системі централізованого тепlopостачання. Апробацію методології виконано на модельному ШІ-дата-центрі потужністю 1 МВт ІТ, розташованому в умовах м. Одеса з локальною PV-станцією. Показано, що оптимальним методом охолодження є пряме рідинне охолодження мікросхем через холодну пластину (direct-to-chip cold plate); рішення є робастним до варіації ваг критеріїв ± 20 %. Підхід розширюється на ШІ-дата-центри більшого масштабу та на нові покоління GPU.

Abstract. A methodology is proposed for selecting the optimal cooling method of an artificial intelligence data center (AI data center) accounting for the integration of renewable energy sources (RES) and smart-grid technologies through a thermal energy storage (TES) system. The developed three-layer decision-making procedure combines a thermodynamic feasibility filter, a posteriori Pareto analysis, and fuzzy multi-criteria convolution within the virtual energy system (VES) architecture. Among the four criteria considered, the smart-grid integration index is of particular significance, being formalized as the present value of two smart-grid services over a five-year horizon: real-time pricing (RTP) arbitrage via TES and waste heat utilization in the district heating system. The methodology is validated on a model AI data center with 1 MW IT capacity, located in Odesa with a local photovoltaic (PV) station. The optimal cooling method is shown to be direct-to-chip cold plate cooling; the decision is robust to ± 20 % variation of criterion weights. The methodology is modular and generalizable to larger AI data centers and new GPU generations.

Keywords: AI data center, cooling methods, immersion cooling, smart grid, renewable energy sources, thermal energy storage, fuzzy multi-criteria optimization, Pareto analysis, virtual energy system.

Ключові слова: ШІ-дата-центр, методи охолодження, занурювальне охолодження, smart grid, відновлювані джерела енергії, теплова акумуляція, нечітка багатокритеріальна оптимізація, Парето-аналіз, віртуальна енергетична система.



Вступ

Швидкий розвиток штучного інтелекту у 2022–2026 рр. суттєво змінив вимоги до фізичної інфраструктури обчислень: енергетичні потреби запитів, що відносяться до ІІІ наближаються до значення на порядок вищого за відповідні запити пошукових систем, а обсяги навчальних навантажень великих мовних моделей подвоюються приблизно кожні 3–4 місяця [5]. Зазначене призвело до появи нового класу інфраструктурних об'єктів — дата-центрів штучного інтелекту (ІІІ-дата-центри), які відрізняються від традиційних і хмарних дата-центрів за питомим тепловим навантаженням на стійку, яке у ІІІ-кластерах становить 50–100 кВт і у 4–7 разів перевищує характерний діапазон конвенційних дата-центрів (5–15 кВт на стійку), субсекундною мінливістю GPU-навантаження та необхідністю взаємодії з електричною мережею як активного споживача.

Сукупність зазначених факторів формує багатокритеріальну задачу вибору методу охолодження ІІІ-дата-центру, що зв'язує термодинаміку охолоджувального циклу, профіль обчислювального навантаження та електричну мережу. У роботі [4] продемонстровано, що холодильні системи в концепції smart grid доцільно розглядати як активні елементи віртуальної енергетичної системи (англ. virtual energy system, VES), здатні виконувати функцію акумуляторів енергії. Загальні принципи інтеграції smart grid з хмарними обчислювальними платформами систематизовано в [11], де обчислювальні навантаження розглядаються як керований ресурс з боку електричної мережі через просторово-часову міграцію задач і використання резервних джерел живлення як елементів demand response. Цей підхід природним чином узагальнюється на дата-центри, де охолоджувальний контур має значну теплову інерцію і може бути доповнений системами теплової акумуляції (TES) [3, 5].

В умовах повоєнного відновлення України, цифрової трансформації та досягнення стійкого розвитку задача набуває ваги: триває розбудова суверенної AI-інфраструктури, зростає частка відновлюваних джерел енергії (ВДЕ, англ. renewable energy sources, RES) в балансі генерації, відбувається перебудова системи централізованого теплопостачання, зберігаються підвищені вимоги до енергонезалежності об'єктів критичної інфраструктури. Метою роботи є розробка процедури вибору оптимального методу охолодження ІІІ-дата-центру з урахуванням інтеграції ВДЕ та smart-grid технологій та її апробація на модельному об'єкті в умовах м. Одеса. Запропонований підхід поєднує методологію нечіткого нелінійного багатокритеріального програмування з тришаровою архітектурою прийняття рішень: фільтр термодинамічної здійсненності - Парето-аналіз ефективності - нечітка згортка переваг.

Аналіз літературних даних і постановка проблеми

Систематичний аналіз енергетичної ефективності дата-центрів закладено у роботах Sun і Lee [1], якими запропоновано метрики, що згодом еволюціонували в показники ефективності використання потужності (PUE) і ефективності використання води (WUE), стандартизовані консорціумом Green Grid. Авторами встановлено, що в обстежених об'єктах система опалення, вентиляції та кондиціонування формує до 52 % сукупного енергоспоживання, причому неоптимальне розміщення обладнання призводить до зростання споживання за фактично нижчого IT-навантаження. В ІІІ-дата-центрах ця проблематика загострюється внаслідок зростання густини теплового потоку та переходу до архітектур з підвищеною теплоємністю. В оглядових роботах [5] детально розглянуто специфіку ІІІ-кластерів: ієрархічну архітектуру систем накопичення енергії, субсекундну варіабельність GPU-навантаження, потребу інтеграції з електричною мережею в режимі активних споживачів — учасників ринку. Зокрема, авторами зазначено, що *системи теплової акумуляції* (TES) є недовикористаним інструментом, який одночасно дозволяє підвищити ефективність охолодження і сприяє інтеграції ВДЕ.

Огляд Kong та ін. [3] систематизує сучасні методи прямого рідинного охолодження дата-центрів: пряме охолодження мікросхем через холодну пластину (direct-to-chip cold plate), занурювальне (immersion), струменеве (spray/jet) та мікроканалне (microchannel). Метод direct-to-chip cold plate передбачає циркуляцію рідкого теплоносія (вода/гліколь) безпосередньо крізь пластину теплообмінника на поверхні процесора або GPU і забезпечує PUE у діапазоні 1,10–1,20 за теплового навантаження до ~150 кВт/стійку. Для занурювального охолодження продемонстровано PUE 1,02–1,10; робочі тіла — діелектричні рідини HFE-7100, Novec 7000, FC-72. Прямі рідинні методи здатні відводити теплові потоки 240–1004 кВт/м² (immersion) і до 17 000 кВт/м² (microchannel), що на два–три порядки перевищує показники повітряного охолодження (CRAC), для якого PUE становить 2,00–2,49 при граничній щільності ~30 кВт/стійку. Щодо капітальних витрат, Kong та ін. [3] оцінюють співвідношення CAPEX immersion-схем до повітряних на рівні до 250 %. Водночас в [3] прямо зазначено дефіцит обґрунтованих керівних принципів вибору методу охолодження та потребу в гібридних підходах до управління.

Задача вибору методу охолодження ІІІ-дата-центру є задачею багатокритеріального прийняття рішень (MCDM). У сучасній теорії виокремлюють апріорні (a priori) методи — зважена сума, метод аналізу ієрархій Сааті [8], метод TOPSIS [9], групи ELECTRE і PROMETHEE — що передбачають артикуляцію переваг через ваги до оптимізації; та апостеріорні (a posteriori) методи, які спочатку формують Парето-фронт ефективних альтернатив без залучення суб'єктивних переваг [10]. Окрему гілку складає теорія нечітких множин Заде [6], що дозволяє формалізувати невизначеність через функції належності, з інтегруванням через адитивну згортку або максимінний оператор Беллмана-Заде [7]. Артеменко і Мазур [4] застосували нечітку нелінійну математичне програмування до холодильних систем у режимі smart grid, де холодильна установка трактується як активний елемент VES. На програмному рівні агент диспетчеризації VES може бути реалізований як сервіс за моделлю



хмарних обчислень у визначенні NIST [15], що забезпечує еластичне масштабування обчислювальних ресурсів при координації регіональних кластерів. Chen та ін. [2] узагальнили просторово-часову регуляцію навантаження дата-центрів та формалізували модель потужності системи охолодження, зручну для зв'язування ІТ-навантаження з електричною мережею. Гібридні підходи, що поєднують апостеріорний Парето-фільтр з апріорною нечіткою згортою, в науковій літературі не набули широкого застосування, попри методологічну перевагу.

Постановка проблеми. Незважаючи на наявність ґрунтовних робіт у кожному напрямі на сьогодні відсутня багатокритеріальна методологія оптимального вибору, яка б одночасно (а) враховувала специфіку ІІІ-дата-центрів через густину теплового потоку та профіль GPU-навантаження; (б) інтегрувала вибір методу охолодження з відновлювальними джерелами енергії та smart-grid через TES в архітектурі VES; (в) використовувала багатопарову процедуру прийняття рішень у конфігурації *термодинамічна здійсненність - Парето-ефективність - нечітка перевага*.

Мета і завдання дослідження

Метою роботи є розробка методології вибору оптимального методу охолодження ІІІ-дата-центру з урахуванням інтеграції відновлюваних джерел енергії та smart-grid технологій, а також моделювання ІІІ-дата-центра в умовах м. Одеса. Для досягнення поставленої мети сформульовано такі завдання: 1) узагальнити вимоги ІІІ-дата-центрів до системи охолодження з огляду на густину теплового потоку GPU-кластерів і профіль обчислювального навантаження; 2) сформулювати порівняльну характеристику термодинамічно здійснених методів охолодження ІІІ-дата-центрів за термодинамічними, енергетичними, економічними та екологічними показниками; 3) розробити тришарову процедуру вибору з фільтром термодинамічної здійсненності, Парето-аналізом ефективності та нечіткою багатокритеріальною згортою у двох варіантах - адитивній [4] та консервативній максимінній [7]; 4) сформулювати модель smart-grid інтеграції методу охолодження з ВДЕ через систему TES в архітектурі віртуальної енергетичної системи; 5) провести апробацію розробленої методології на модельному ІІІ-дата-центрі потужністю 1 МВт ІТ в м. Одеса з модельним RTP-тарифом, базовими вагами критеріїв $w = (0,30; 0,30; 0,20; 0,20)$ та провести аналіз чутливості.

Методи і матеріали досліджень

Модельна конфігурація ІІІ-дата-центру. Об'єктом дослідження є модельний ІІІ-дата-центр потужністю 1 МВт ІТ у м. Одеса. Конфігурація: 64 сервери DGX H100 (8 × NVIDIA H100 SXM5, TDP 700 Вт/GPU; сумарна потужність сервера ~10,2 кВт з урахуванням CPU, пам'яті та допоміжних підсистем), 8 стійок по 8 серверів із середньою потужністю ~82 кВт на стійку з урахуванням мережевого обладнання, мережа InfiniBand NDR 400 Гбіт/с. Електропостачання здійснюється від мережі 35/0,4 кВ спільно з локальною сонячною електростанцією потужністю 1 МВт-пік та системою акумуляування електричної енергії (BESS) ємністю 0,5 МВт-год. Теплова підсистема включає систему теплової акумуляції (TES) ємністю 2 МВт-год та підключення до магістралі централізованого теплопостачання. Вибір параметрів модельного об'єкта 1 МВт ІТ відповідає практичній нижній межі для комерційного GPU-кластера рівня DGX H100; безпосередньо на GPU припадає ~358 кВт (512 GPU × 700 Вт), решта — на CPU, пам'ять, NVSwitch, мережеву фабрику та системи зберігання. Потужність сонячної електростанції 1 МВт-пік відповідає площі ~0,5 га. Ємність TES 2 МВт-год забезпечує пікове зміщення до 2 годин повного охолоджувального навантаження, що є типовим для комерційних дата-центрів [3]. Ємність BESS 0,5 МВт-год відповідає UPS-рівню резервування ~30 хв при повному ІТ-навантаженні. Структурну схему енергетичних потоків модельного ІІІ-дата-центру наведено на рис. 1.

Термодинамічна модель методів охолодження. Множина методів: $M = \{M_1, M_2, M_3\}$ визначається як {CRAC, direct-to-chip cold plate, immersion}. Оскільки питома щільність ІТ-навантаження 62 кВт/стійку перевищує практичну межу повітряного охолодження ~30 кВт/стійку [3], метод M_1 відхиляється фільтром здійсненності (рівн. 5), і термодинамічна модель будується для $M' = \{M_2, M_3\}$. Для обох методів теплове навантаження від ІТ-обладнання становить $Q_i = P_{IT}$. Для M_2 тепловіддача здійснюється через одноконтурний рідинний тракт (вода/гліколь), для M_3 – через двофазне кипіння діелектричної рідини HFE-7100 з конденсацією пари у вторинному контурі. Динамічний енергобаланс охолоджувального контуру методу m [4]:

$$M_m C_{pm} \frac{dT_m}{dt} = Q_a + Q_i - Q_e; \quad Q_a = U A_a (T_a - T_m) \quad (1)$$

де M_m - теплова маса робочого тіла в контурі методу m , кг; C_{pm} - питома теплоємність, Дж/(кг·К); T_m - температура робочого тіла, °С; T_a - температура навколишнього середовища, °С; U - коефіцієнт теплопередачі, Вт/(м²·К); A_a - ефективна площа теплообміну, м²; Q_a - конвективний теплообмін з навколишнім середовищем, Вт; $Q_i = P_{IT}$ — теплове навантаження від ІТ-обладнання, Вт; Q_e - потужність, відведена системою охолодження (паро-компресійний цикл або вільне охолодження), Вт.

Паро-компресійний цикл, що активується при $T_a > T_{fc}$ (температурний поріг перемикавання), описується рівняннями [4]:

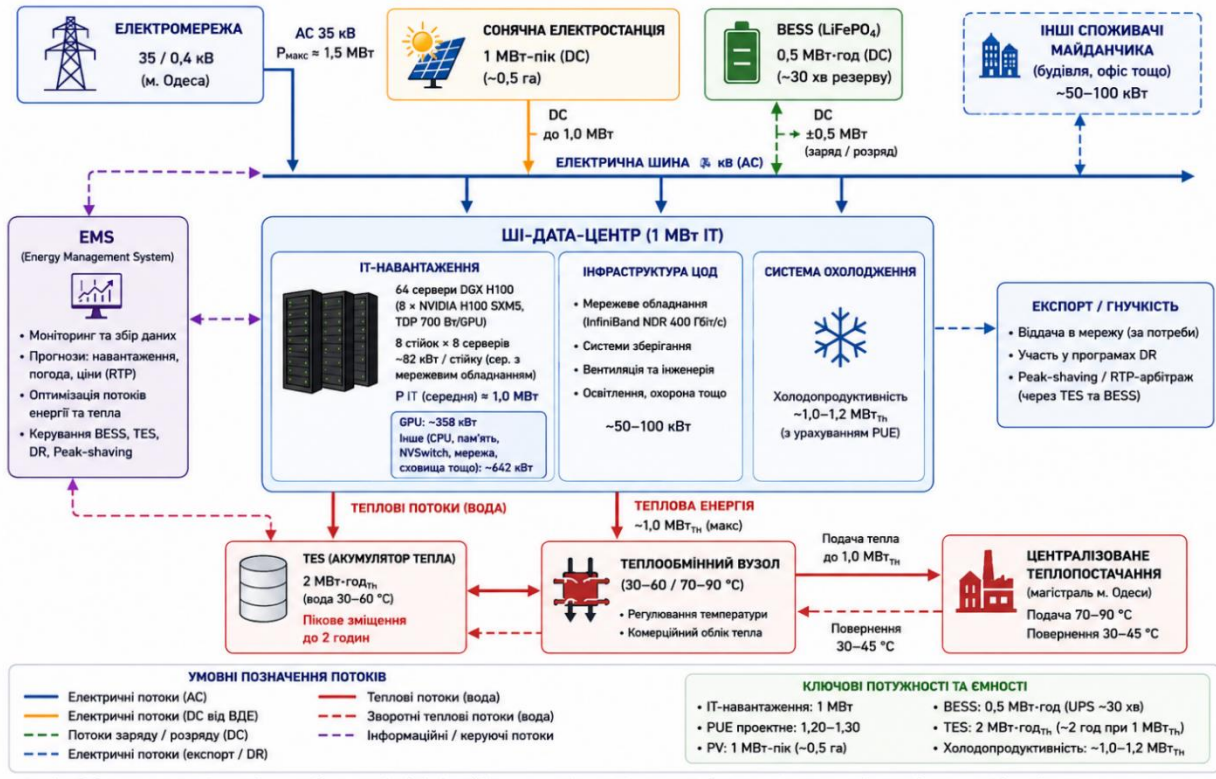
$$Q_e = \dot{m}_{comp}(h_1 - h_4); \quad Q_c = \dot{m}_{comp}(h_2 - h_3); \quad W_{comp} = \frac{\dot{m}_{comp}(h_{2s} - h_1)}{\eta_s}; \quad h_4 = h_3 \quad (2)$$

де \dot{m}_{comp} - масова витрата холодоагенту, кг/с; h_1-h_4 - питомі ентальпії холодоагенту у вузлових точках циклу: вхід компресора, вихід компресора, вихід конденсатора, вихід розширювального клапана, відповідно, Дж/кг;



h_{2s} - ентальпія після ізентропного стиснення, Дж/кг; η_s — ізентропний ККД компресора; Q_c - теплота, відведена у конденсаторі, Вт; W_{comp} - потужність компресора, Вт. Умова $h_4 = h_3$ відповідає ізентальпійному процесу в розширювальному клапані.

Рис. 1. Структурна схема енергетичних потоків модельного ШІ-дата-центру (1 МВт ІТ, м. Одеса)



Примітки: 1) Значення потужностей наведені як середні/максимальні робочі оцінки. 2) Напрямки потоків показані для типової роботи; можливі двонаправлені режими (експорт, заряд/розряд).

Рис. 1. Структурна схема енергетичних потоків модельного ШІ-дата-центру
Fig. 1. Block diagram of energy flows in a model AI data center

Термодинамічні властивості робочих тіл (вода/гліколь для M_2 , HFE-7100 для M_3) вибрані з NIST Reference Fluid Database [12] через бібліотеку CoolProp [13]. Для умов м. Одеса взяті ТМУ-дані, $T_{fc} = 18$ °C. Потужність системи охолодження апроксимується лінеаризованою моделлю [2]:

$$P_{cool}(t) = (PUE_m - 1) \cdot P_{IT}(t) \quad (3)$$

де PUE_m — ефективність використання потужності методу m : $PUE_{M_2} \approx 1,15$; $PUE_{M_3} \approx 1,07$.

Управління TES реалізовано як задача диспетчеризації в агентній архітектурі, де агент щогодини визначає розклад заряду та розряду TES, розв'язуючи задачу мінімізації річних витрат на електроенергію з мережі за тарифом реального часу (RTP)

$$\min \sum_{t=1}^{8760} c_t \cdot [P_{cool}(t) - P_{RES}(t) - P_{TES}^{dis}(t)]^+ \quad (4)$$

де c_t — погодинний RTP-тариф; $P_{RES}(t)$ — генерація СЕС відповідно до ERA5 та PVGIS [14].

Нечітка процедура вибору охоплює шість кроків.

Крок 1. Фільтр здійсненності. Метод m вважається здійсненим, якщо максимальна питома потужність охолодження $q_{max}(m)$ не менша за потребу об'єкта q_{req} :

$$F(m) = \begin{cases} 1, & q_{max}(m) \geq q_{req} \\ 0, & \text{інакше} \end{cases} \quad (5)$$

Оскільки $q_{req} = 62$ кВт/стійку, а для повітряного охолодження $q_{max}(M_1) \approx 30$ кВт/стійку, маємо $F(M_1) = 0$; подальша процедура застосовується до $M' = \{M_2, M_3\}$.

Крок 2. Критерії оцінювання. Для кожного $m \in M'$ обчислюються натуральні значення чотирьох критеріїв $c_k(m)$, $k \in \{1,2,3,4\}$: C_1 — річне споживання електроенергії з мережі, МВт·год/рік (мінімізувати); C_2 — приведена п'ятирічна вартість охолоджувальної підсистеми (CAPEX + OPEX), млн USD (мінімізувати); C_3 — нормований екологічний слід (агрегат CO₂-емісій та WUE), в.о. (мінімізувати); C_4 — приведена п'ятирічна вартість smart-grid сервісів, тис. USD (максимізувати):

$$C_4(m) = AF \cdot [E_{RTP}(m) + E_{heat}(m)] \quad (6)$$



де $AF = \sum_{t=1}^5 (1+r)^{-t}$ — ануїтетний коефіцієнт (при $r = 5\%$: $AF = 4,329$). Компонента RTP-арбітражу через TES:

$$E_{RTP}(m) = N_{cycles} \cdot E_{TES} \cdot [RTE(m) \cdot (c_{peak} - c_{off}) - (1 - RTE(m)) \cdot c_{off}] \quad (7)$$

де N_{cycles} - кількість повних циклів заряд/розряд TES за рік; E_{TES} - ємність TES, МВт·год; $RTE(m)$ - round-trip efficiency: $RTE_{M_2} = 0,85$, $RTE_{M_3} = 0,70$; c_{peak} , c_{off} — піковий і нічний тарифи, USD/МВт·год. Компонента утилізації скидної теплоти системи централізованого теплопостачання:

$$E_{heat}(m) = \dot{Q}_{cool}(m) \cdot \eta_{rec}(m) \cdot h_{heat} \cdot c_{heat} \quad (8)$$

де $\dot{Q}_{cool}(m)$ - потужність охолодження, МВт; $\eta_{rec}(m)$ - коефіцієнт рекуперації скидної теплоти: $\eta_{rec,M_2} = 0,80$, $\eta_{rec,M_3} = 0,20$; h_{heat} - тривалість опалювального сезону, год/рік; c_{heat} - тариф на теплову енергію, USD/МВт·год.

Крок 3. Парето-фільтр. Метод m_i домінує m_j (позначається $m_i > m_j$), якщо $c_k(m_i) \leq c_k(m_j)$ для $k \in \{1,2,3\}$ та $c_4(m_i) \geq c_4(m_j)$, причому хоча б одна нерівність має бути строгою. Парето-ефективна множина:

$$P = \{m \in M' : \nexists m' \in M', m' > m\}$$

Крок 4. Функції належності. Лінійні функції належності будуються за абсолютними референтними діапазонами $[c_k^{min}, c_k^{max}]$, визначеними незалежно від складу M' :

$$\mu_k(m) = \frac{c_k^{max} - c_k(m)}{c_k^{max} - c_k^{min}} \quad \text{для } C_1, C_2, C_3 \quad (9)$$

$$\mu_k(m) = \frac{c_k(m) - c_k^{min}}{c_k^{max} - c_k^{min}} \quad \text{для } C_4 \quad (10)$$

Використання абсолютних діапазонів забезпечує незалежність функцій належності від кількості альтернатив і дозволяє коректно порівнювати методи в різних сценаріях без повторної нормалізації.

Крок 5. Нечіткі згортки. Для підвищення робастності застосовуються дві незалежні схеми:

$$S^{add}(m) = F(m) \cdot \sum_{k=1}^4 w_k \mu_k(m), \quad \sum_{k=1}^4 w_k = 1 \quad (11)$$

$$S^{BZ}(m) = F(m) \cdot \min_{k \in \{1..4\}} \mu_k(m) \quad (12)$$

де w_k - ваги критеріїв; S^{add} - адитивна згортка [4]; S^{BZ} - максимінний оператор Беллмана–Заде [7], що гарантує відсутність компенсації слабкого критерію за рахунок сильних. Якщо обидві схеми дають однаковий m^* , рішення вважається робастним.

Крок 6. Рішення. $m^* = \operatorname{argmax}_{m \in P} S(m)$. Базові ваги: $w = (0,30; 0,30; 0,20; 0,20)$. Аналіз чутливості виконується варіацією w_k у межах $\pm 20\%$ від базових значень та альтернативними сценаріями розвитку електричної мережі.

Моделювання реалізовано за допомогою Python (NumPy, Pandas, CoolProp) з годинною дискретизацією (8760 точок). ТМУ-дані для м. Одеса: середньорічна температура $+10,5^\circ\text{C}$, понад 5400 год нижче $+18^\circ\text{C}$. ВДЕ-генерація за PVGIS [14] і ERA5: СЕС 1296 МВт·год/рік ($CF = 14,8\%$). RTP-тариф апроксимовано двозонним розподілом 60–110 USD/МВт·год. Емісійний коефіцієнт: 290 г $\text{CO}_2/\text{кВт}\cdot\text{год}$.

Значення критеріїв M наведено в таблиці 1. CAPEX розраховано за галузевими референсами [3] з air-cooling baseline 2,0 млн USD/МВт IT та коефіцієнтами 1,70 (M_2) і 2,30 (M_3). Розрахунок C_4 за рівняннями (6)–(8): 88,5 тис USD для M_2 і 27,9 тис USD для M_3 за 5 років.

Таблиця 1 - Значення критеріїв для методів M_2 і M_3

Критерій	Розм.	M_2	M_3	Перевага
C_1 - споживання з мережі	МВт·год/рік	8 778	8 077	M_3
C_2 - 5-річна вартість, у т. ч.:	млн USD	4,32	5,47	M_2
- CAPEX	млн USD	3,40	4,60	-
- OPEX електроенергії	млн USD	0,48	0,23	-
- обслуговування	млн USD	0,44	0,60	-
- робочі рідини	млн USD	-	0,04	-
C_3 - екол. слід ($\text{CO}_2 + \text{WUE}$)	в. о.	0,553	0,348	M_3
C_4 - 5-річні smart-grid серв., у т. ч.:	тис USD	88,5	27,9	M_2
- RTP-арбітраж (TES)	тис USD	39,8	20,4	-
- утилізація теплоти (ЦП)	тис USD	48,7	7,6	-

Метод M_3 має перевагу за споживанням з мережі (на 8,7 %) і екологічним слідом — внаслідок нижчого PUE. Натомість M_2 виграє за вартістю (на 21 %) та за критерієм smart-grid інтеграції у понад 3 рази.

Жоден з методів не домінує інший: M_2 кращий за C_2 і C_4 , M_3 — за C_1 і C_3 . Парето-фронт $P = \{M_2, M_3\}$. Питома вартість зекономленої електроенергії при переході $M_2 \rightarrow M_3$ складає 1640 USD/МВт·год за 5 років — майже у чотири рази перевищує середній RTP-тариф 85 USD/МВт·год, що вказує на економічну невиправданість переходу до immersion-схеми без значущих додаткових мотивацій.



Функції належності та нечіткі згортки. Референсні діапазони: C_1 [7096; 21900] МВт·год/рік; C_2 [2,5; 8,0] млн USD; C_3 [0; 1]; C_4 [0; 200] тис. USD. Результати наведено в таблиці 2.

Таблиця 2 – Функції належності та значення згортки

Показник	M_2	M_3
μ_1	0,886	0,934
μ_2	0,669	0,460
μ_3	0,447	0,652
μ_4	0,443	0,140

Обидві схеми одностайно вказують $m^* = M_2$. Значення $S^{BZ}(M_3) = 0,140$ обмежене знизу критерієм μ_4 , що сигналізує про структурний дефект immersion-схеми у площині smart-grid інтеграції. Природа цього обмеження фізична: низькотемпературний рівень теплоти конденсації діелектричної рідини (~50–60 °C) обмежує пряму подачу теплоносія в системи ЦТП, що потребують подавальної температури 80–110 °C; як наслідок, коефіцієнт рекуперації · теплоти вже в моделі складає лише 0,20 проти 0,80 для cold plate.

Аналіз чутливості наведено в таблиці 3. Рішення $m^* = M_2$ зберігається у всіх практично релевантних сценаріях.

Таблиця 3 — Чутливість рішення m^* до сценаріїв

Сценарій	w_1	w_2	w_3	w_4	$S^{add}(M_2) / S^{add}(M_3)$	m^*
Базовий	0,30	0,30	0,20	0,20	0,644 / 0,576	M_2
Пріоритет енергії	0,50	0,20	0,20	0,10	0,710 / 0,703	M_2
Пріоритет вартості	0,20	0,50	0,20	0,10	0,645 / 0,561	M_2
Пріоритет екології	0,20	0,20	0,40	0,20	0,579 / 0,568	M_2
Пріоритет smart-grid	0,20	0,20	0,20	0,40	0,577 / 0,465	M_2
Екстремальна екологія	0,10	0,10	0,70	0,10	0,513 / 0,609	M_3
Подвоєний RTP-коридор	0,30	0,30	0,20	0,20	0,695 / 0,618	M_2

Прийняте рішення $m^* = M_2$ зберігається у всіх практично релевантних сценаріях. Інверсія на користь M_3 відбувається лише при $w_3 \geq 0,60$ — сценарій скоріше регуляторно-нормативний, ніж економічний. Сценарій з подвоєним RTP-коридором посилює перевагу M_2 через зростання вартості C_4 .

Обговорення результатів

Інтерпретація основного результату. Перевага методу M_2 над M_3 обумовлена поєднанням незалежних факторів. По-перше, CAPEX-премія immersion-схеми становить ~35 % (4,60 проти 3,40 млн USD), що за поточних капітальних вартостей перевищує дисконтований ефект економії електроенергії за 5-річний горизонт. По-друге, помірне значення PUE для cold plate (1,15) досить близьке до immersion (1,07), щоб різниця у річному споживанні не компенсувала вартісну премію. Таким чином, рішення $m^* = M_2$ є наслідком структури економічних і термодинамічних параметрів задачі. Критерій C_4 , що відповідає за smart-grid складову виявився найбільш дискримінантним показником.

Найбільшим джерелом залишкової невизначеності є точність CAPEX-оцінок. Співвідношення вартості immersion до повітряного базису варіює 200–300 % залежно від виробника та конфігурації. Зниження CAPEX immersion на 20 % або зростання CAPEX cold plate на 10 % скорочує маржу адитивної згортки, але не інвертує рішення. Серед спрощень моделі: постійне IT-навантаження 1 МВт упродовж року; двозонний RTP-профіль замість повної годинної кривої; неявне врахування витрат на підключення до систем централізованого теплопостачання через коефіцієнт рекуперації; обмеження масштабом 1 МВт. Для III-дата-центрів 5–10+ МВт ефект економії на масштабі знижуватиме питомий CAPEX immersion на 10–15 %, що зсуне баланс на користь M_3 .

На світовому ринку у 2024–2026 рр. домінує застосування саме прямого рідинного охолодження мікросхем через холодну пластину як де-факто стандарту для GPU-щільностей 50–80 кВт/стіжку (NVIDIA HGX H100, H200, GB200-NVL72), тоді як immersion залишається нішевим рішенням [3, 5]. Це узгоджується з висновками нашої моделі. З масовим переходом на GPU класу B200/B300 (TDP 1000–1200 Вт) питома щільність стійки досягне 100–130 кВт і вище. За таких умов cold plate наблизатиметься до своїх практичних меж (~150 кВт/стіжку), а immersion отримає природну термодинамічну перевагу. Прогнозований часовий горизонт цього зсуву — 2027–2030 рр.

Як напрями подальших досліджень вбачаються чотири логічних напрями розвитку: (1) застосування алгоритму NSGA-II [10] для одночасної оптимізації проектних параметрів усередині кожного методу (ємність TES, BESS, потужність PV); (2) параметричне дослідження чутливості рішення до покоління GPU (H100 → B200 → R-series); (3) інтеграція з повноцінними програмами demand response Ukrenergo для дата-центрів масштабу 5+ МВт; (4) багатооб'єктна координація регіональних кластерів III-дата-центрів в архітектурі агентної моделі VES [4].

Висновки

Розроблено тришарову процедуру вибору методу охолодження III-дата-центру, яка поєднує методологію нечіткого нелінійного багатокритеріального програмування [4] з апостеріорним Парето-фільтром і забезпечує робастність рішення до варіації преференцій оператора. Запропоновано формальне визначення smart-grid критерію C_4 через дві фізично інтерпретовані компоненти — RTP-арбітраж через TES та утилізацію скидної



теплоти в системі централізованого теплопостачання, що робить цей критерій співмірним з вартісним C_2 і незалежно верифікованим за галузевими даними.

На модельному ШІ-дата-центрі потужністю 1 МВт IT із 512 GPU NVIDIA H100 в умовах м. Одеса встановлено, що оптимальним методом охолодження є пряме рідинне охолодження мікросхем через холодну пластину (direct-to-chip cold plate). Перевага підтверджена обома схемами згортки: адитивною 0,644 проти 0,576 та максимінною 0,443 проти 0,140.

Рішення є робастним до варіації ваг критеріїв $\pm 20\%$ від базових значень $w = (0,30; 0,30; 0,20; 0,20)$, до подвоєного RTP-коридору та до прогнозованого зростання частки ВДЕ в системі генерації. Інверсія на користь immersion-схеми відбувається лише при екстремальних розкладах преференцій ($w_3 \geq 0,60$), що відповідає скоріше регуляторно-нормативним, ніж економічним мотивам. Запропонована методологія є модулярною, припускає масштабування на ШІ-дата-центри більшого масштабу та нові покоління GPU, що становить предмет подальших досліджень із залученням алгоритму NSGA-II.

Список використаних джерел

1. Sun H. S., Lee S. E. Case study of data centers' energy performance. *Energy and Buildings*. 2006. Vol. 38, No. 1. P. 522–533. DOI: 10.1016/j.enbuild.2005.08.012.
2. Chen M., Gao C., Song M., Chen S., Li D., Liu Q. Internet data centers participating in demand response: A comprehensive review. *Renewable and Sustainable Energy Reviews*. 2020. Vol. 117. Article 109466. DOI: 10.1016/j.rser.2019.109466.
3. Kong R., Zhang H., Tang M., Zou H., Tian C., Ding T. Enhancing data center cooling efficiency and ability: A comprehensive review of direct liquid cooling technologies. *Energy*. 2024. Vol. 308. Article 132846. DOI: 10.1016/j.energy.2024.132846.
4. Artemenko S., Mazur V. Cloud-computing and Smart Grid for Energy-Efficient Refrigeration Systems. Artemenko S., Mazur V. Cloud-computing and Smart Grid for Energy-Efficient Refrigeration Systems. *Proceedings of IEEE Conference*. 2023. [вказіть назву конференції, DOI або місце проведення].
5. Ginzburg-Ganz E., Lifshits P., Machlev R., Belikov J., Krieger Z., Levron Y. Technical challenges of AI data center integration into power grids — a survey. *Energies*. 2025. Vol. 19, No. 1. Article 137.
6. Zadeh L. A. Fuzzy sets. *Information and Control*. 1965. Vol. 8, No. 3. P. 338–353.
7. Bellman R. E., Zadeh L. A. Decision-making in a fuzzy environment. *Management Science*. 1970. Vol. 17, No. 4. P. B141–B164.
8. Saaty T. L. *The Analytic Hierarchy Process*. New York : McGraw-Hill, 1980. 287 p.
9. Hwang C. L., Yoon K. *Multiple Attribute Decision Making: Methods and Applications*. Berlin : Springer-Verlag, 1981. 259 p.
10. Deb K., Pratap A., Agarwal S., Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*. 2002. Vol. 6, No. 2. P. 182–197.
11. Markovic D., et al. Smart power grid and cloud computing. *Renewable and Sustainable Energy Reviews*. 2013. Vol. 24. P. 566–577.
12. NIST Reference Fluid Thermodynamic and Transport Properties Database. URL: <https://webbook.nist.gov> (дата звернення: 26.05.2026).
13. Bell I. H., Wronski J., Quoiloin S., Lemort V. Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library CoolProp. *Industrial & Engineering Chemistry Research*. 2014. Vol. 53, No. 6. P. 2498–2508.
14. Photovoltaic Geographical Information System (PVGIS). European Commission Joint Research Centre. URL: https://re.jrc.ec.europa.eu/pvg_tools/en/ (дата звернення: 26.05.2026).
15. Mell P., Grance T. *The NIST Definition of Cloud Computing*. NIST Special Publication 800-145. 2011. 7 p.

Отримана в редакції 19.01.2026. Прийнята до друку 26.01.2026. Розміщено в інтернеті 30 березня 2026.

Received 19 January 2026. Approved 26 January 2026. Available in Internet 30 March 2026