

УДК 004:697.1:725.1

DOI: 10.15587/1729-4061.2018.147485

Підготовка та початковий аналіз даних щодо енергоспоживання муніципальних будівель

А. Л. Перекрест, О. П. Чорний, О. В. Мур, А. В. Ніколенко, В. В. Кузнецов,
Є. В. Кузнецова

Систематизація даних щодо енергоспоживання будівель різного призначення дозволяє дослідити процеси з позицій ефективного використання енергоресурсів для забезпечення комфортних умов. Це обумовлює необхідність удосконалення існуючих або пошуку нових підходів до аналізу даних щодо енергоспоживання різних будівель.

На основі дослідження процесу підготовки даних про енергоспоживання будівель запропоновано методика первинного аналізу. Враховуються призначення будівлі та способи отримання даних, інформація щодо показників абсолютного й питомого електро- та теплоспоживання, показники температур внутрішнього і зовнішнього повітря. На прикладі енергоспоживання будівель навчального закладу виконана апробація розробленої методики початкового аналізу даних.

Дослідження енергоспоживання будівель навчального закладу дозволили встановити кількісну зміну енерговитрат та температур зовнішнього й внутрішнього повітря при переході від річних даних до даних опалювального періоду. Аналіз діаграм розмаху наглядно демонструють діапазони зміни показників енергоспоживання, їх середні значення та викиди, що вказують на збільшену понад норму витрату енергоресурсів досліджуваними об'єктами в перехідні періоди.

Для наочної оцінки ефективності використання одиниці теплової енергії одиницею площі різних квартир з різними способами обліку та регулювання теплоспоживання розроблено web-базоване програмне рішення, у якому порівнюються показники теплоспоживання окремих квартир з максимальною потребою на опалення відповідно до нормативних документів.

Отримані результати слугують основою для створення прикладних інформаційних рішень для муніципального енергоменеджменту

Ключові слова: теплове забезпечення будівель, аналіз даних енергоспоживання, підготовка даних, моніторинг енергоспоживання будівель

1. Вступ

Сьогодні значна увага приділяється проблемам ефективного використання енергоресурсів та енергозбереження. Заходи, що здійснюються з метою енергозбереження, зазвичай не відрізняються значною ефективністю. Це пов'язано з відсутністю повної інформації про об'єкти енергоспоживання та специфіку

експлуатації, відсутністю необхідної кількості коштів на проведення заходів, спрямованих на підвищення рівня енергоефективності будівель тощо.

Рішення проблеми енергоефективності передбачає облік енергоресурсів, аналіз режимів енергоспоживання, прийняття рішень і керування. Практична реалізація, як правило, основана на постійному моніторингу та параметризації системи, що спричиняє накопичення великого обсягу історичних даних про енергоспоживання, що отримані від вузлів комерційного обліку та електронних погодозалежних регуляторів.

Рішення проблем енергоефективності та енергозбереження потребує комплексного підходу, що обумовлює необхідність удосконалення існуючих або пошуку нових підходів до аналізу даних щодо енергоспоживання [1–3]. Обробка значних об'ємів даних обумовлює використання спеціалізованих алгоритмів та методів статистичного та інтелектуального аналізів [4–6].

Збір даних, підготовка та проведення початкового аналізу дозволяє отримати цінну інформацію щодо енергоспоживання до створення моделі методами машинного навчання. Початковий аналіз значно впливає на адекватність моделей, що створюються. Об'єктивними є труднощі, пов'язані з специфікою предметної області, поєднанням даних з декількох джерел та відсутності уніфікованої методики аналізу даних, результати використання якої можна б було імплементувати в існуючі системи.

Розвиток інформаційних технологій призвів до зростання кількості сфер застосування методів машинного навчання. Серед них – управління енергоспоживанням та прогнозування витрат енергоресурсу. Методи машинного навчання допомагають в аналізі параметрів системи управління енергоспоживанням та підтримці прийняття ефективних рішень енергоменеджерами.

Існує необхідність удосконалення наявних або пошуку нових підходів до аналізу даних енергоспоживання, з метою прийняття рішень, спрямованих на підвищення ефективності використання енергоресурсів.

Впровадження рішень щодо автоматизованого моніторингу та контролю систем опалення дозволяють зменшувати загальне споживання теплоенергії та накопичувати великі обсяги інформації про функціонування системи та прийнятті енергоменеджером рішення. Вирішуються проблеми, пов'язані з необхідністю забезпечення нормативних санітарно-гігієнічних умов у приміщеннях, що опалюються, з урахуванням впливу зовнішніх факторів.

2. Аналіз літературних даних та постановка проблеми

У роботі [7] розглядається можливість прогнозування теплоспоживання в системі централізованого тепlopостачання, що базується на використанні нейронних мереж Ельмана з використанням генетичного алгоритму, який дозволяє оптимізувати ваги моделей. Дослідження [8] підтверджує, що результати, отримані за допомогою нейронних мереж, відрізняються більшою точністю в порівнянні з ансамблевим методом *random forest* у задачі прогнозування споживання електроенергії. Проте обидві моделі можуть використовуватися для годинного прогнозування. Авторами [9] проаналізовано характеристики енергоспоживання й фактори, що впливають на нього, на базі даних про функціону-

вання будівель магазинів. Оскільки використовувалися різноманітні методи інтелектуального аналізу значного об'єму даних, отримані висновки дозволяють оптимізувати витрати на будівництво та бізнес-обладнання на основі створеної прогнозуючої моделі.

Незважаючи на практичну значущість таких результатів, не розглянуто в достатній мірі процес імплементації отриманих результатів в існуючі системи підтримки прийняття рішень. Наведений опис власне процесу аналізу не є достатньо задокументованим. Крім того, розроблений лише для дослідження даних енергоспоживання магазинів, він передбачає використання специфічних факторів його оцінки. Дослідженнями точності моделей з використанням штучних нейронних мереж для короткострокового прогнозування теплозабезпечення об'єктів соціально-бюджетної сфери встановлено, що найкращих результатів досягає модель типу NARX [10]. Оскільки дані, що досліджуються, зазвичай, змішаного типу, неправильно структуровані та/або мають пропущені значення в роботі [11] синтезовано моделі імпутації даних на основі декількох методів машинного навчання. Проведено порівняння з типовим методом, що включений у більшість статистичних пакетів.

В роботі [12] представлено метод діагностики енергетичної ефективності на основі обмеженої кількості даних з рахунків за електроенергію та даних про погодні умови. Отримані рішення забезпечують можливість на основі результатів багатопараметричної регресії визначати кількість спожитого ресурсу кожною з систем, а також детальну інформацію про фізичні властивості. Також запропоновано шляхи отримання необхідних даних. Тим не менш, не дивлячись на переваги методу, залишається відкритим питання підготовки даних та попереднього аналізу.

Розвиток поглядів на дослідження та, власне, набори даних про енергоспоживання може бути знайдений в роботі [13]. У ній проводиться класифікація існуючих наборів даних відповідно до трьох визначених стратегій генерування та розробка каталогу доступних даних про енергоспоживання. Автори роботи обмежуються лише збором даних та їх групуванням, а також загальним представленням класифікаційного апарату.

Дослідники недостатньо уваги приділяють питанням збору даних, підготовки та проведення початкового аналізу, що дозволить отримати інформацію щодо енергоспоживання до створення моделі та значно вплине на адекватність моделі. Отже, існує необхідність у створенні уніфікованої методики дослідження даних щодо енергоспоживання будівель різного функціонального призначення.

3. Ціль та задачі дослідження

Проведені дослідження ставили за мету підвищення якості аналізу енергоспоживання муніципальних будівель за рахунок автоматизації процесу первинної обробки даних, що були отримані з комерційних вузлів обліку чи встановлених систем моніторингу енергоспоживання.

Для досягнення поставленої мети необхідно вирішити наступні задачі:

– визначити особливості підходів до дослідження даних щодо енергоспоживання, адаптувати існуючі методології відповідно до особливостей предметної області та створити методику початкового аналізу даних щодо енергоспоживання будівель;

– обґрунтувати структури інформаційних потоків даних про енергоспоживання різних будівель, реалізувати комплекс програмних рішень для отримання даних з різних джерел при забезпеченні необхідного рівня їх інформативності та якості та провести апробацію методики первинного аналізу даних на даних енергоспоживання будівель навчального закладу;

– розробити web-базоване, прикладне програмне рішення для наглядного аналізу витрат теплової енергії різними квартирами з різними способами обліку та регулювання теплоспоживання.

4. Матеріали та методи дослідження даних

4. 1. Інформаційні потоки даних

Галуззю, що споживає значну частину енергоресурсів, є теплопостачання об'єктів житлово-комунального господарства міст: житловий фонд, будинки соціального, навчального призначення тощо.

Об'єкти, дані з яких досліджувалися, умовно розділено на чотири групи: перша група – це будинки з тепловим лічильником, друга – будинки без теплового лічильника, третя – приватні будинки, четверта – адміністративні будівлі.

В якості адміністративних будівель було обрано навчальні корпуси Кременчуцького національного університету ім. Михайла Остроградського (КрНУ, Україна). У КрНУ розроблена і встановлена автоматизована система для оперативного контролю та управління (АСМУ) тепловими пунктами окремих навчальних корпусів. Система містить стандартне промислове обладнання. Завдяки розробленому програмному забезпеченню здійснюється поточний контроль температурних режимів систем опалення і оперативну зміну уставок температури теплоносія в залежності від зовнішньої температури і умов роботи корпусів [14]. В результаті роботи АСМУ теплоспоживанням навчальних корпусів стають доступними дані про реальні режими роботи систем опалення та енерговитрати окремих будівель.

Об'єктивним та доступним джерелом для отримання початкових даних будівель без встановленої САР є рахунки на оплату комунальних послуг.

Для адміністративних і бюджетних будівель відомі витрати теплоенергії за лічильником будинку, електроенергії за лічильником будинку, зовнішній об'єм будівель та вартість теплової енергії.

Аналіз рахунків показав, що для будинків з тепловим лічильником доступні дані про площу квартири, вартість теплової енергії, частки квартири від опалювальної площі будинку, витрати за лічильником квартири, будинку, нарахована сума.

Аналогічно для квартир в будинках без теплових лічильників доступними є площа квартири, тариф та нарахована сума на оплату послуги.

Аналіз рахунків приватних будинків показав, що відомими даними є об'єм спожитого газу, опалювальна площа, вартість газу.

4. 2. Методика початкового аналізу даних

Data Mining – процес, мета якого полягає в знаходженні значущих кореляцій, моделей та трендів шляхом обробки великого об'єму даних з використанням різноманітних статистичних і математичних методів. Типовим є використання міжгалузевого стандарту дослідження даних CRISP-DM. Згідно з ним життєвий цикл має складатися з шести фаз: розуміння бізнес-процесів, розуміння даних, підготовка даних, моделювання, оцінка, впровадження.

При аналізі даних щодо енергоспоживання існує необхідність вирішувати певні бізнес-завдання: Які будівлі є найбільш енергоефективними? Які будівлі є найбільш комфортними для перебування людей? Які будівлі є найбільш інвестиційно привабливими? Які будівлі потребують першочергової модернізації? Наскільки якісно надається послуга з централізованого опалення?. Тому, доцільно в дослідженні даних керуватися однією з існуючих методологій для аналізу. Враховуючи, що аналітична обробка даних енергоспоживання є типовим проектом Data Mining, на основі CRISP-DM з урахуванням специфіки предметної області, вирішено створити уніфіковану методику початкового аналізу даних щодо енергоспоживання.

Створена методика передбачає наступні етапи:

- збір вихідних даних: запит на доступ до даних зі списку ресурсів проекту і їх вивантаження;
- опис даних: дослідження основних або «поверхневих» властивостей зібраних даних і опис результатів: обсяг даних, типи значень, схеми кодування;
- дослідження даних: аналіз «сирих» даних за допомогою запитів, візуалізації і звітності;
- перевірка якості даних: наявність помилок кодування та відсутніх значень;
- вибір даних: вибір елементів та атрибутів;
- проведення ряду синтаксичних перетворень, викликаних специфікою функціонування обраного інструментарію для аналізу;
- відбір та об'єднання даних: злиття таблиць даних, які містять різну або однакову інформацію про одні й ті ж об'єкти;
- породження ознак: сумарне енергоспоживання, питоме тепло-, електро- та сумарне енергоспоживання, створені ознаки, приведені до нормативних значень зовнішньої та внутрішньої температури повітря, кількість градусо-днів, мітка опалювального сезону (за необхідності);
- аналіз загальних описових статистик для набору даних, що досліджується на даному етапі;
- перевірка на нормальність розподілу: проведення тесту Шапіро-Уїлка, за необхідності проведення окремого дослідження асиметрії та ексцесу з урахуванням розміру вибірки;
- проведення кореляційного аналізу для кількісної оцінки взаємозв'язку між даними;
- побудова діаграм розмаху (boxplot) або ядерної оцінки густини для порівняння розподілу ймовірностей між групами даних та аналіз.

Відповідно до створеної методики наводяться такі описові статистики: середнє арифметичне, середнє квадратичне відхилення, медіана, мінімум, максимум, асиметрія, ексцес, стандартна похибка середнього.

Для перевірки на нормальність розподілу проводиться тест Шапіро-Уїлка. Нульова гіпотеза, що перевіряється за допомогою тесту Шапіро-Уїлка, формулюється так: «Вибірка, що аналізується, походить з генеральної сукупності, що має нормальний розподіл». Якщо отримана за допомогою тесту ймовірність помилки виявляється меншою від заздалегідь прийнятого рівня значущості (наприклад, 0,05), нульова гіпотеза відхиляється.

Зважаючи на результати перевірки на нормальність розподілу, визначається метод кореляції: коефіцієнт Пірсона – для нормально розподілених даних, непараметричний коефіцієнт Спірмена – для даних з розподілом, що відрізняється від нормального.

5. Результати досліджень енергоспоживання будівель

5. 1. Структури інформаційних потоків даних про енергоспоживання різних будівель

Для спрощення та зручності у підготовці даних розроблено структури інформаційних потоків для різних об'єктів.

Так, для будинку з тепловим лічильником (рис. 1) спочатку визначають загальну опалювальну площу:

$$S_{\text{заг}} = \frac{S_{\text{кв}} \cdot 100 \%}{dS_{\text{кв}}}, \quad (1)$$

де $S_{\text{заг}}$ – загальна опалювальна площа, м^2 ; $S_{\text{кв}}$ – площа квартири, м^2 ; $dS_{\text{кв}}$ – частка квартири від загальноопалюваної площі.

Потім визначається питоме теплоспоживання:

$$q = \frac{\Delta E_{\text{буд}} \cdot 1163}{S_{\text{буд}}}, \quad (2)$$

де q – питоме теплоспоживання будинку, $\text{кВт} \cdot \text{год} / \text{м}^2$; $S_{\text{буд}}$ – площа будинку, м^2 ; $\Delta E_{\text{буд}}$ – витрати за лічильником, Гкал.

Клас енергоефективності:

$$K_{ee} = \frac{q - E_{\text{max}}}{E_{\text{max}}}, \quad (3)$$

де K_{ee} – коефіцієнт енергоефективності; E_{max} – максимальна питома потреба на опалення, $\text{кВт} \cdot \text{год} / \text{м}^2$; q – питоме теплоспоживання, $\text{кВт} \cdot \text{год} / \text{м}^2$.

Тариф на опалення:

$$T = \frac{\Delta E_{\text{буд}} \cdot B_{\text{те}}}{S_{\text{заг}}}, \quad (4)$$

де T – тариф на опалення, грн/м²; $B_{\text{те}}$ – вартість теплової енергії, грн/Гкал; $S_{\text{заг}}$ – загальна опалювана площа, м²; $\Delta E_{\text{буд}}$ – витрати за лічильником, Гкал.

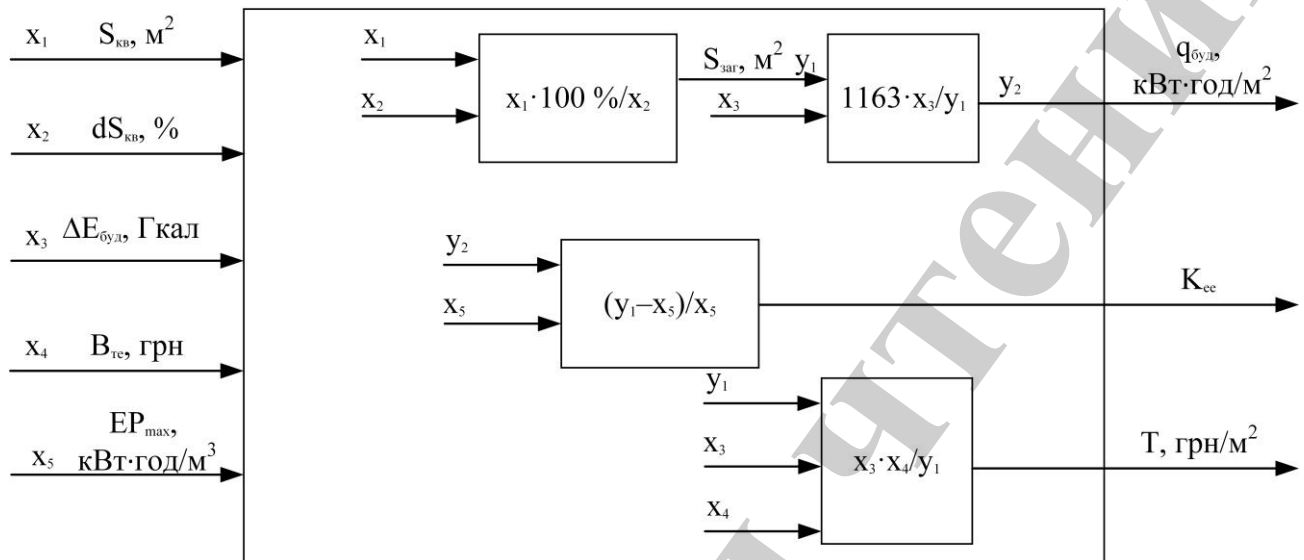


Рис. 1. Структура інформаційних потоків для будинків з тепловим лічильником

Для будинків без теплового лічильника (рис. 2) визначається кількість теплової енергії на опалення приміщень:

$$\Delta E_{\text{кв}} = \frac{H_{\text{кв}}}{B_{\text{те}}}, \quad (5)$$

де $\Delta E_{\text{кв}}$ – кількість теплової енергії, використаної для опалення приміщення, Гкал; $H_{\text{кв}}$ – нарахована сума, грн; $B_{\text{те}}$ – вартість теплової енергії, грн/Гкал.

Наступними етапами є визначення питомого теплоспоживання (2) та коефіцієнта енергоефективності (3).

Оскільки загальна опалювальна площа невідома, то показники визначаються для квартири (рис. 3).

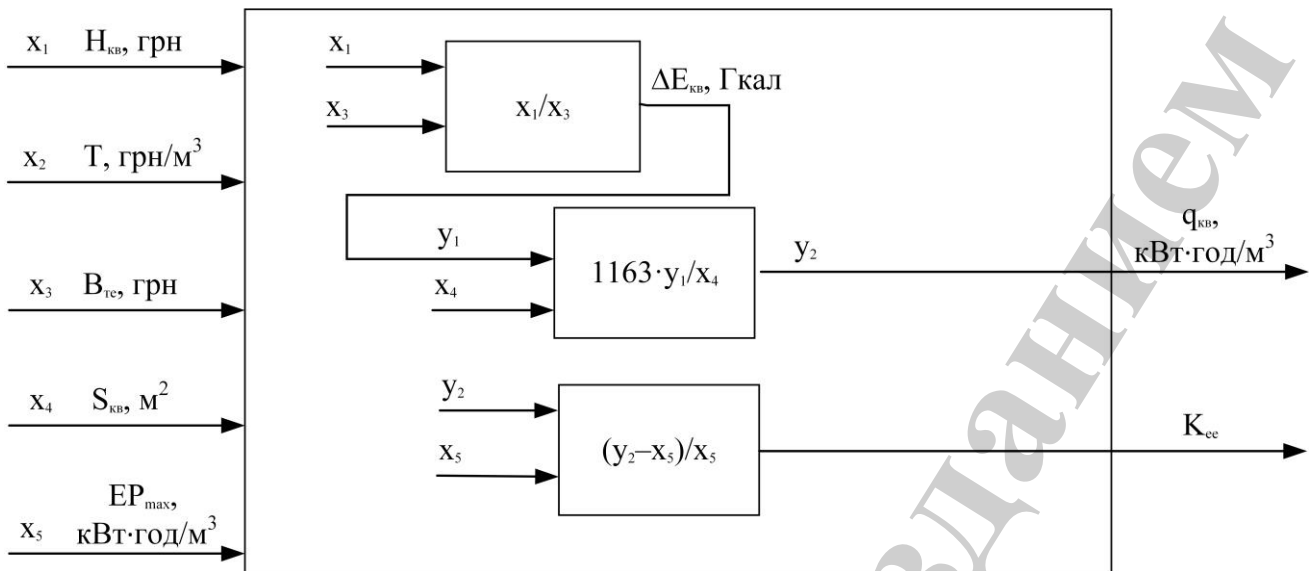


Рис. 2. Структура інформаційних потоків для будинків без теплового лічильника

Для визначення показників теплоспоживання приватних будинків з лічильником природного газу (рис. 3) спочатку початкові дані необхідно привести до стандартних умов, помноживши спожитий об'єм газу його на відповідний коефіцієнт:

$$V_c = V_r \cdot K_i, \quad (6)$$

де V_c – об'єм газу приведений до стандартних умов, m^3 ; V_r – об'єм газу у робочих умовах, m^3 ; K_i – коефіцієнт приведення газу до стандартних умов.

Використання того чи інакшого коефіцієнта залежить від місця розміщення газового лічильника, регіону або області та відповідного календарного місяця. Оскільки лічильник природного газу фіксує загальний об'єм спожитого газу абонентами, то виникає задача визначення того об'єму газу, що використовується саме на опалення будинку:

$$V_{ci} = V_c \cdot V_{поб}, \quad (7)$$

де $V_{поб}$ – об'єм газу, що використовується на побутові потреби. V_{ci} – об'єм газу, використаний на опалення, m^3 ; V_c – об'єм газу, приведений до стандартних умов, m^3 .

Потім визначаються абсолютні теплові витрати на опалення приватного будинку:

$$\Delta E_{буд} = \frac{T_{зг} \cdot V_{ci}}{10^6}, \quad (8)$$

де $\Delta E_{\text{буд}}$ – абсолютні теплові витрати на опалення будинку, Гкал; $T_{\text{зг}}$ – теплота згоряння газу (калорійність), ккал/м³; V_{ci} – об’єм газу, використаний саме на опалення, м³.

Калорійність є основним показником якості природного газу. Чим вища калорійність, тим менший об’єм газу необхідний для задоволення певних потреб. Значення калорійності фіксуються постачальниками послуг та зазвичай є доступними для використання.

Тариф на опалення:

$$T = \frac{V_{ci} \cdot B_{\Gamma}}{S_{\text{буд}}}, \quad (9)$$

де T – тариф на опалення грн/м²; B_{Γ} – вартість газу, грн/м³; $S_{\text{буд}}$ – площа будинку, м²; V_{ci} – об’єм газу, використаний саме на опалення, м³.

Отримані результати дозволяють визначити питоме теплоспоживання та клас енергоефективності (2–3).

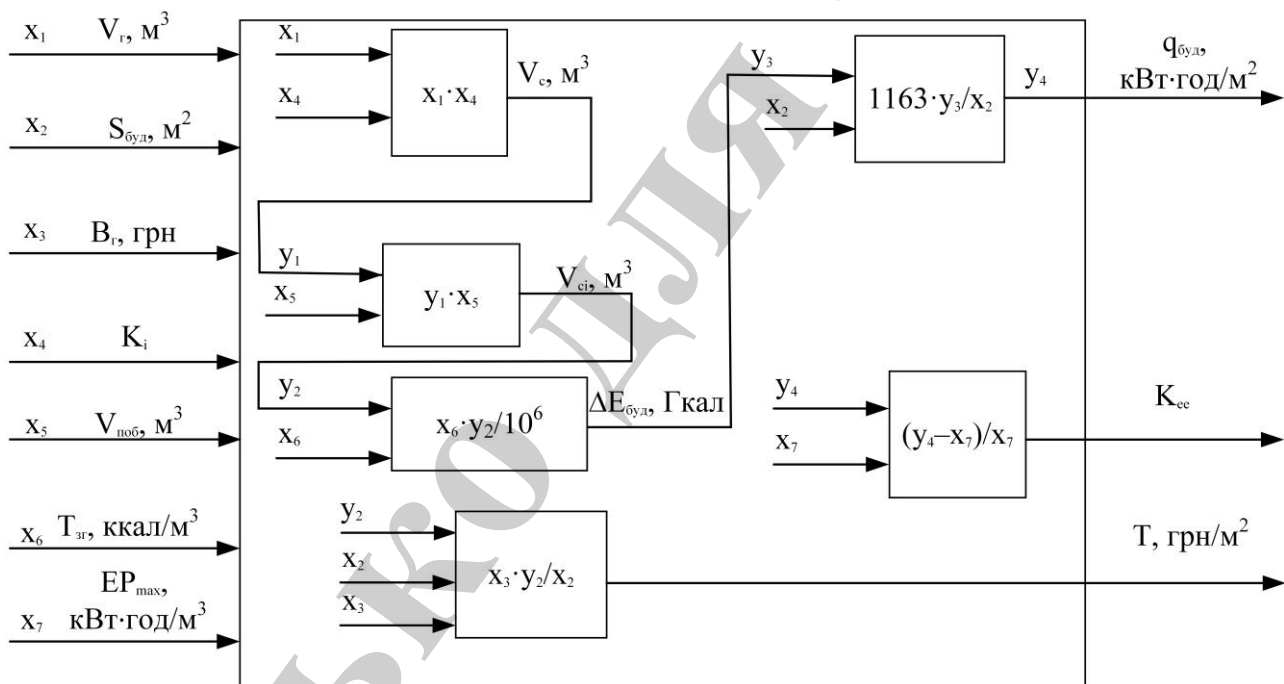


Рис. 3. Структура інформаційних потоків для приватних будинків

Для адміністративних і бюджетних будівель (рис. 4) проводиться розрахунок, аналогічний будівлям з тепловим лічильником (2–4), та отримуються значення питомого теплоспоживання, тарифу та класу енергоефективності.

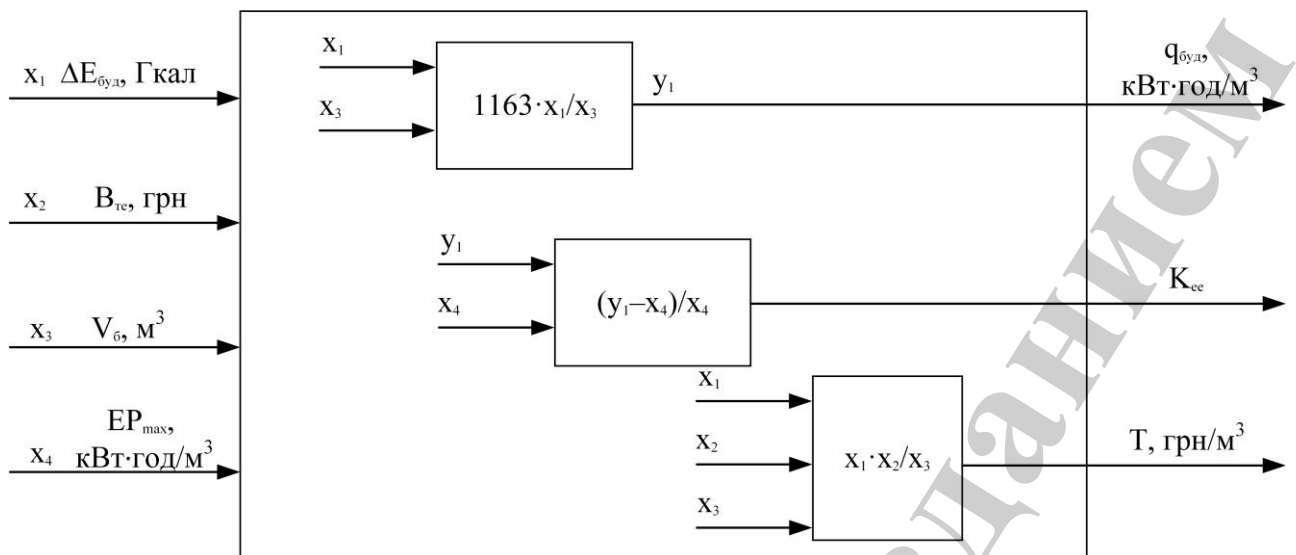


Рис. 4. Структура інформаційних потоків для адміністративних та бюджетних будівель

5. 2 Апробація методики первинного аналізу даних

Для апробації створеної методики у якості даних були вибрані дані щодо енергоспоживання семи навчальних корпусів КрНУ за період з 2012 по 2016 роки.

Для даного дослідження було зібрано три набори даних:

– Енергоспоживання будівель навчального закладу за період з 2012 по 2016 роки із зазначенням номера корпусу, місяця, року, споживання теплоенергії (вираженого в Гкал), споживання електроенергії (вираженого в кВт·год) (табл. 1).

Кількість даних: стовпців – 5, рядків – 420.

Формат даних: номер корпусу – категоріальна змінна, решта даних – числові.

Таблиця 1

Схема кодування набору «Енергоспоживання та внутрішня температура»

Номер корпусу	k
місяць	month
рік	year
Теплоспоживання, Гкал	E_Gkal
Електроспоживання, кВт·год	W_kWh
Середньомісячна внутрішня температура, °C	Tin_C

– Обсяги теплового навантаження, із зазначенням номера корпусу, обсягів теплового навантаження (виражених в Гкал/год) та об'єму будівлі (табл. 2).

Кількість даних: стовпців – 3, рядків – 7.

Формат даних: номер корпусу – категоріальна змінна, решта даних – числові.

– Середньомісячна температура навколишнього середовища за період з 2012 по 2016 роки, із зазначенням року, місяця та середньомісячної температури (табл. 3).

Кількість даних: стовпців – 3, рядків – 60.

Формат даних: всі дані – числові.

Таблиця 2

Схема кодування набору «Обсяги теплового навантаження»

Номер корпусу	k
Об'єм, m^3	V_m3
Обсяг теплового навантаження, Гкал/год	Qop

Таблиця 3

Схема кодування набору «Температура»

Рік	$year$
Місяць	$month$
Середньомісячна зовнішня температура, $^{\circ}C$	$Tout_C$

Початкове дослідження обмежиться даними за 2012–2016 роки (420 елементів), тому фільтри потрібно налаштувати таким чином, щоб дані за інші періоди не опинилися у вибраних даних.

Набори даних, що досліджуються, не містять помилок та пропущених значень, а схеми кодувань збігаються. Після перевірки якості переходимо до фази підготовки даних.

Усі зазначені вище атрибути були включені в датасет. Оскільки початкові зібрані дані про енергоспоживання представлені у першій нормальній формі, здійснено перехід до третьої нормальної форми, для усунення транзитивних залежностей. Результатом проведення об'єднання трьох таблиць є єдина таблиця з помісячними даними кожного корпусу за 2012–2016 роки, що містить атрибути, вибрані на попередніх кроках аналізу.

Для оцінки початкових даних виникає необхідність у додаткових параметрах для порівняльного аналізу різних об'єктів. Для порівняльного аналізу енергоспоживання різних будівель доцільно використовувати однакові показники, наприклад, питоме теплоспоживання ($q1$, $kWh \cdot год / m^3$), питоме електроспоживання ($q2$, $kWh \cdot год / m^3$) тощо.

Нормативне значення градусо-днів опалювального періоду (ГДОП):

$$Dd_H = (T_{вн} - T_{з_ср_н}) \cdot Z_H, \quad (10)$$

де Dd_H – нормована кількість градусо-днів; $T_{вн}=20$, $^{\circ}C$ – внутрішня температура приміщення; $T_{з_ср_н}=-0,8$, $^{\circ}C$ – середня нормована зовнішня температура; $Z_H=180$ – нормативна тривалість опалювального періоду.

Фактичне значення градусо-днів опалювального періоду (ГДОП):

$$Dd = (T_{in_C} - T_{out_C}) \cdot Z_{\text{факт}}, \quad (11)$$

де Dd – фактична кількість градусо-днів; T_{in_C} , °C – внутрішня температура приміщення; T_{out_C} , °C – середня фактична зовнішня температура; $Z_{\text{факт}}$ – фактична тривалість опалювального періоду.

Коефіцієнт ГДОП:

$$K_{Dd} = \frac{Dd_n}{Dd}, \quad (12)$$

де K_{Dd} – коефіцієнт ГДОП; Dd – фактична кількість градусо-днів; Dd_n – нормована кількість градусо-днів.

Абсолютне теплоспоживання у кВт·год:

$$E_kWh = E_Gkal \cdot 1163, \quad (13)$$

де E_kWh – абсолютне теплоспоживання, кВт·год; E_Gkal – витрати теплової енергії, Гкал.

Питоме теплоспоживання:

$$q1 = \frac{E_kWh}{V}, \quad (14)$$

де $q1$ – питоме теплоспоживання, кВт·год/м³; E_kWh – витрати теплової енергії, кВт·год; V – об'єм будівлі, м³.

Питоме електроспоживання:

$$q2 = \frac{W_kWh}{V}, \quad (15)$$

де $q2$ – питоме електроспоживання, кВт·год/м³; W_kWh – витрати електричної енергії, кВт·год; V – об'єм будівлі, м³.

Сумарне енергоспоживання:

$$E_{\text{sum_kWh}} = E_kWh + W_kWh, \quad (16)$$

де $E_{\text{sum_kWh}}$ – сумарне енергоспоживання, кВт·год; W_kWh – витрати електричної енергії, кВт·год; E_kWh – витрати теплової енергії, кВт·год.

Питоме сумарного енергоспоживання:

$$q3 = \frac{E_{sum_kWh}}{V}, \quad (17)$$

де $q3$ – питома сумарне енергоспоживання, кВт·год/м³; E_{sum_kWh} – питома енергоспоживання, кВт·год; V – об'єм будівлі, м³.

Питома теплоспоживання, приведене до нормативних значень зовнішньої та внутрішньої температури повітря:

$$q1t = q1 \cdot K_{Dd}, \quad (18)$$

де $q1t$ – питома теплоспоживання з урахуванням температур, кВт·год/м³; $q1$ – питома теплоспоживання, кВт·год/м³; K_{Dd} – коефіцієнт ГДОП.

Питома електроспоживання, приведене до нормативних значень зовнішньої та внутрішньої температури повітря:

$$q2t = q2 \cdot K_{Dd}, \quad (19)$$

де $q2t$ – питома електроспоживання з урахуванням температур, кВт·год/м³; $q2$ – питома електроспоживання, кВт·год/м³; K_{Dd} – коефіцієнт ГДОП.

Питома сумарне енергоспоживання, приведене до нормативних значень зовнішньої та внутрішньої температури повітря:

$$q3t = q3 \cdot K_{Dd}, \quad (20)$$

де $q3t$ – питома сумарне енергоспоживання з урахуванням температур, кВт·год/м³; $q3$ – питома сумарне енергоспоживання, кВт·год/м³; K_{Dd} – коефіцієнт ГДОП.

У випадку аналізу енергоспоживання будівель, при переході від загальних даних до даних опалювальних та неопалювальних періодів, відбувається перетворення вибірки даних споживання, де кожен запис відповідає місячним показникам, в нову вибірку, де кожному корпусу вже відповідає ряд відсортованих записів із характеристиками по окремо взятому опалювальному (неопалювальному) періоду. Для аналізу посезонного енергоспоживання створюється єдина нова ознака – мітка опалювального сезону.

Після проводиться перетворення таблиці з даними місячного енергоспоживання будівель навчального закладу. де кожен запис відповідає одному місяцю, в нову таблицю. У ній кожному споживачеві енергії вже відповідає один запис із загальними характеристиками по його енергоспоживанню (наприклад, загальна кількість спожитої електроенергії за опалювальної період, кількість градусо-днів цього періоду тощо).

На рис. 6 наведені графіки матриць кореляцій загальних даних (рис. 5), помісячних даних опалювальних (рис. 6) і неопалювальних сезонів (рис. 7) та сумарних даних опалювальних (рис. 8) і неопалювальних сезонів (рис. 9).

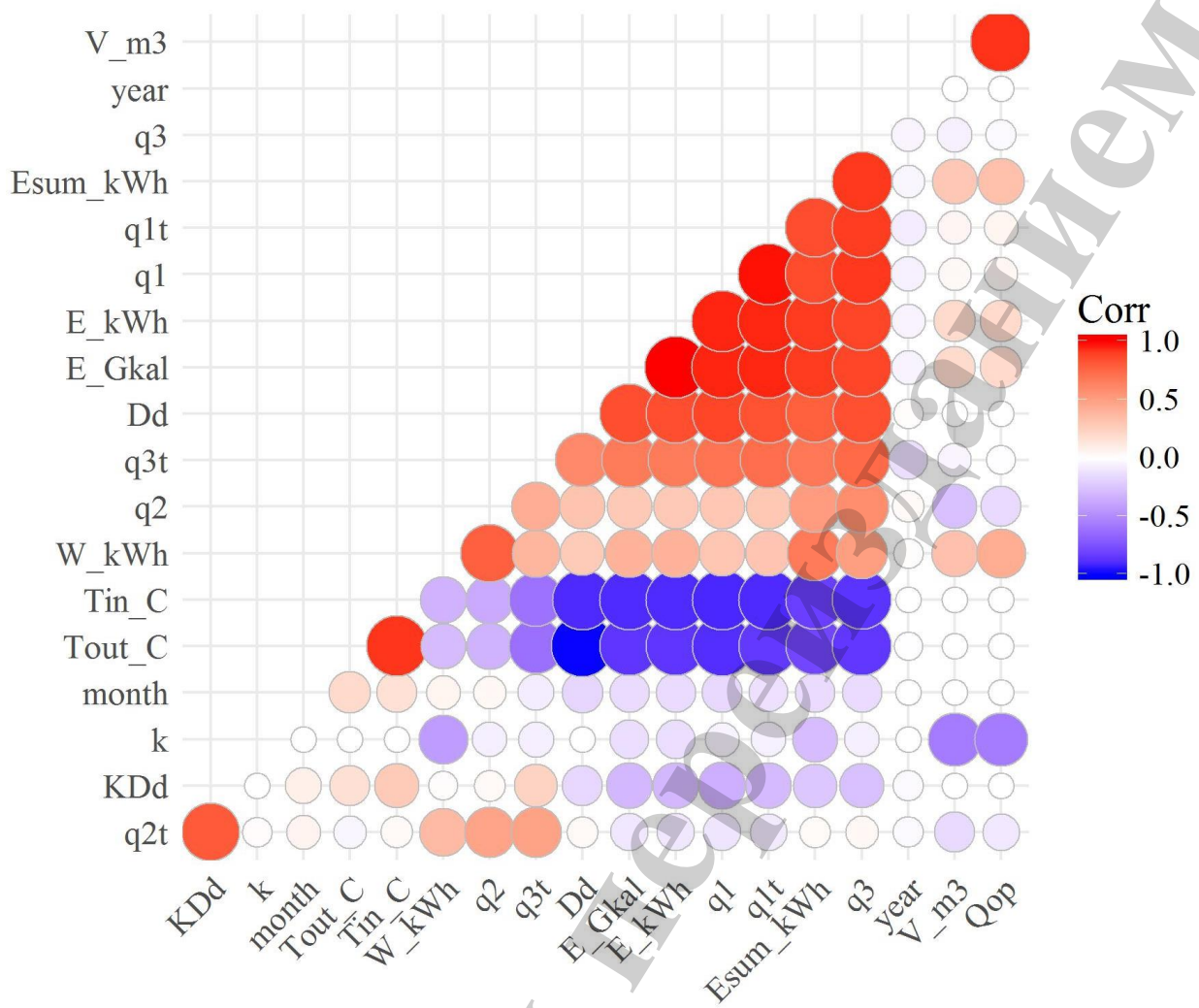


Рис. 5. Матриця кореляцій помісячних даних (без поділу на опалювальні та неопалювальні сезони) про енергоспоживання

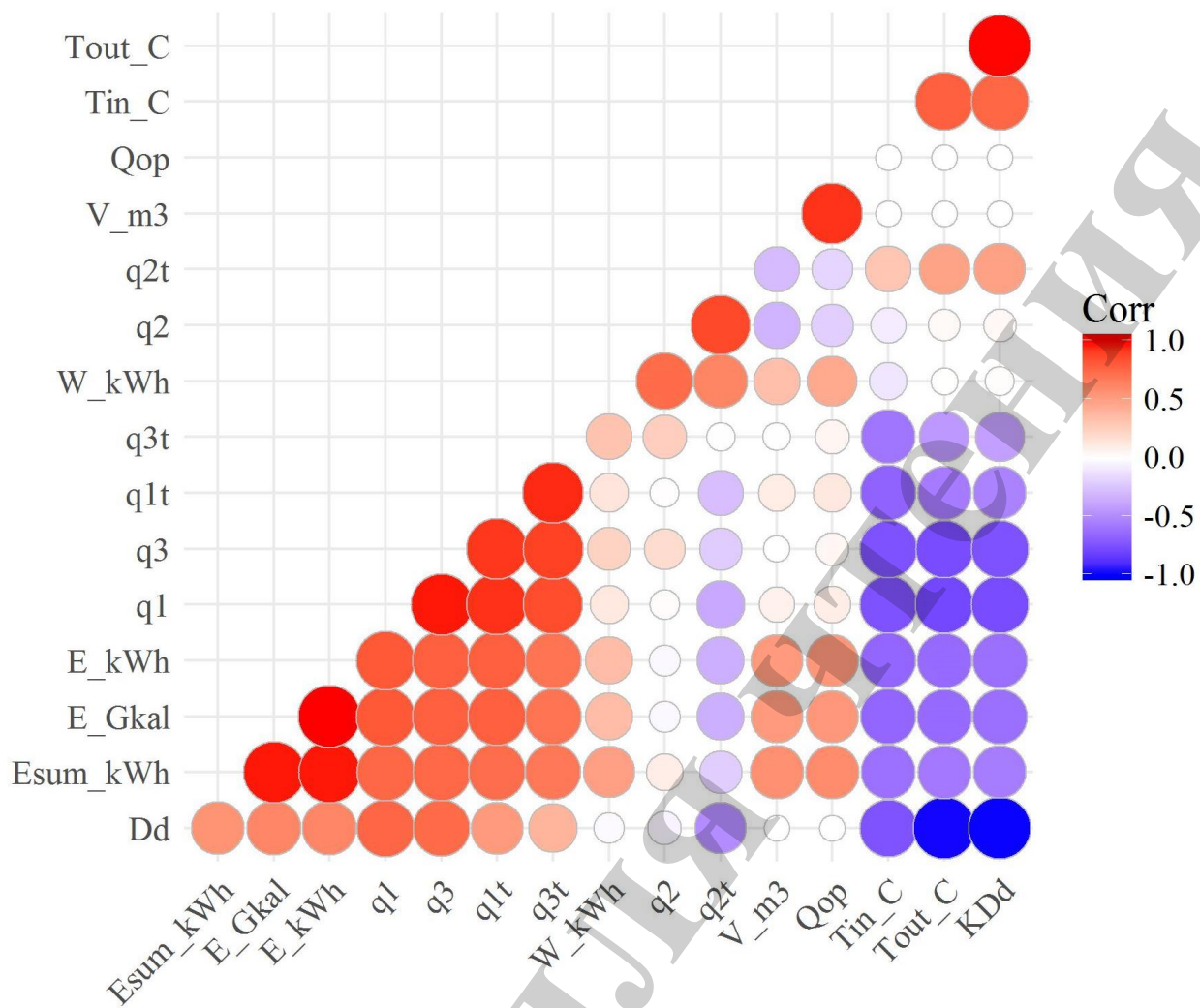


Рис. 6. Матриця кореляцій помісячних даних опалювальних сезонів про енергоспоживання

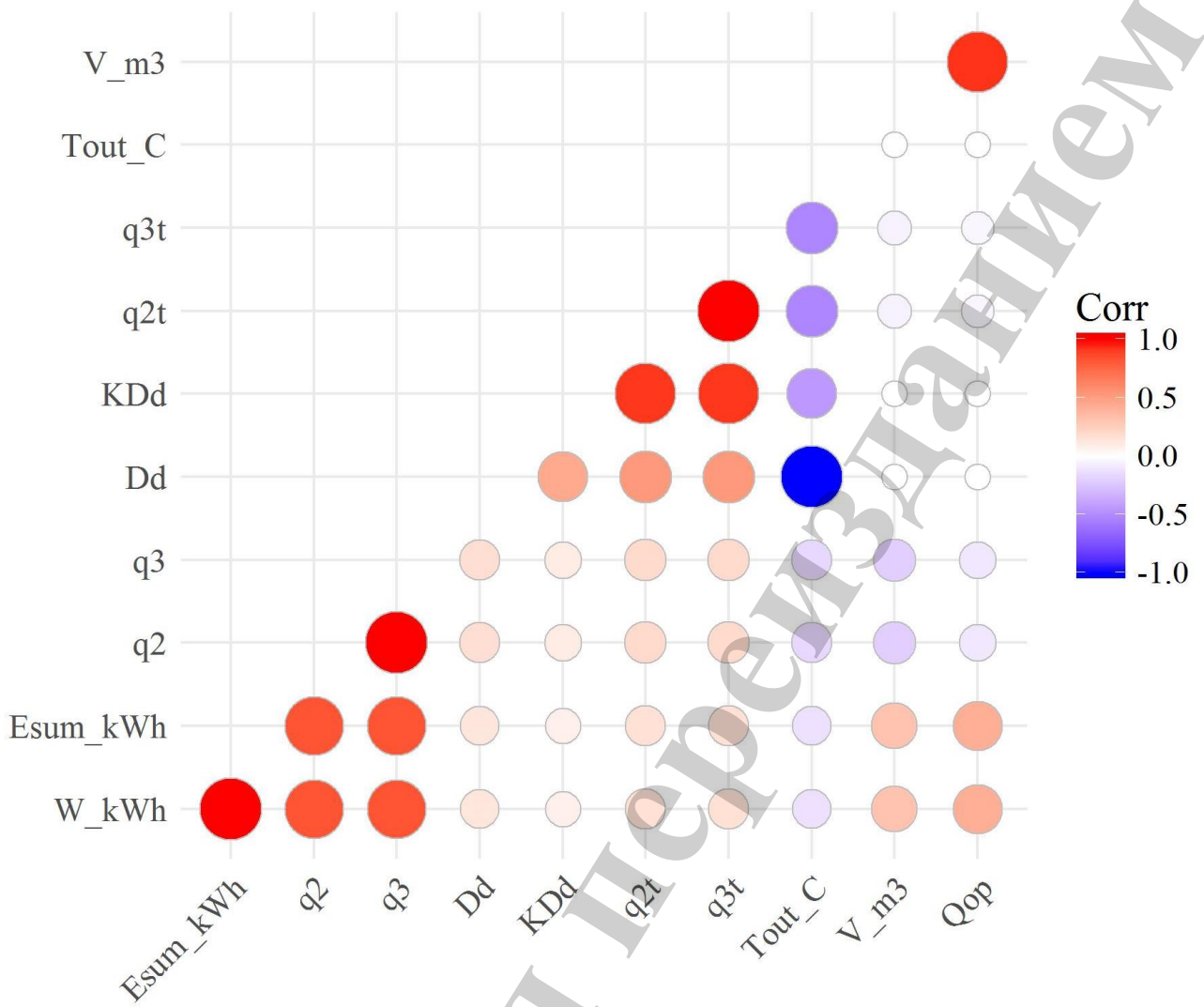


Рис. 7. Матриця кореляцій помісячних даних неопалювальних сезонів про енергоспоживання

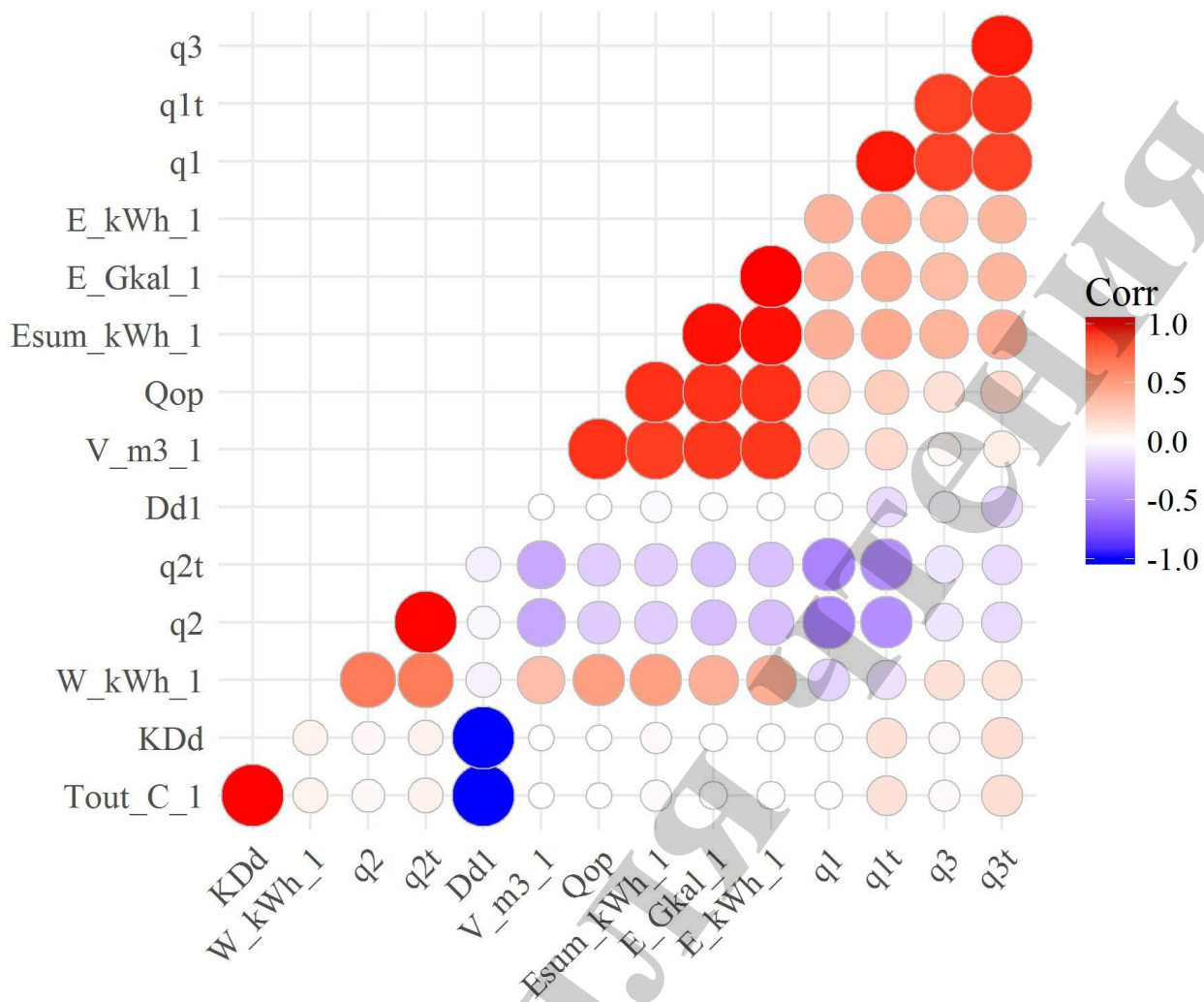


Рис. 8. Матриця кореляцій сумарних даних опалювальних сезонів про енергоспоживання

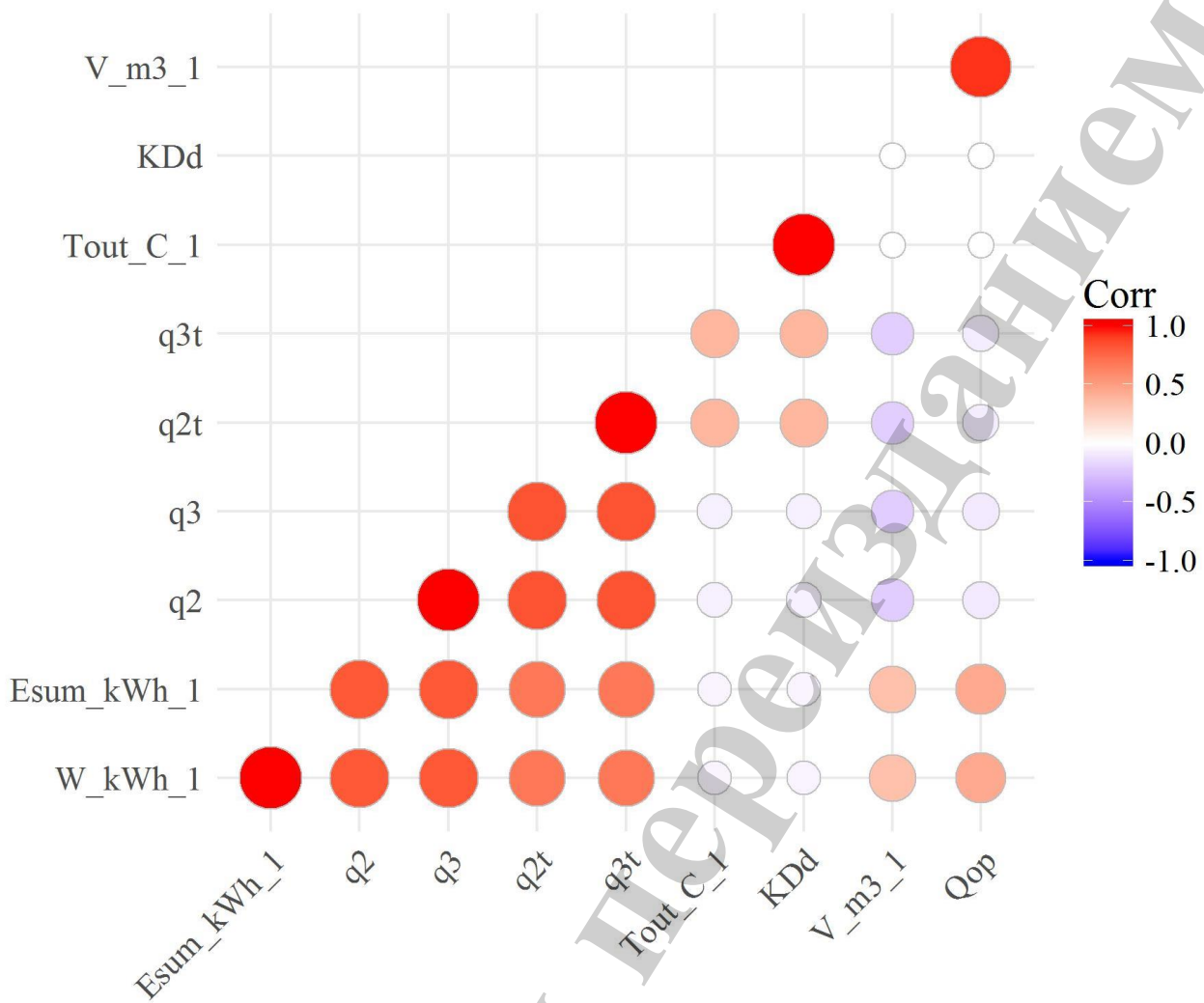


Рис. 9. Матриця кореляцій сумарних даних неопалювальних сезонів про енергоспоживання

Наступним етапом є аналіз діаграм розмаху. На рис. 10–15 наведені графіки для посезонного енергоспоживання за опалювальні сезони (теплоспоживання та сумарне енергоспоживання).

Проаналізовано дані щодо теплоспоживання. В опалювальному сезоні 2012–2013 абсолютне теплоспоживання п'ятого корпусу (при тепловому навантаженні, що дорівнює 0,2761 Гкал/год) перевищило той же показник другого корпусу (при тепловому навантаженні, що дорівнює 0,3019 Гкал/год) на 102568 кВт·год (рис. 10).

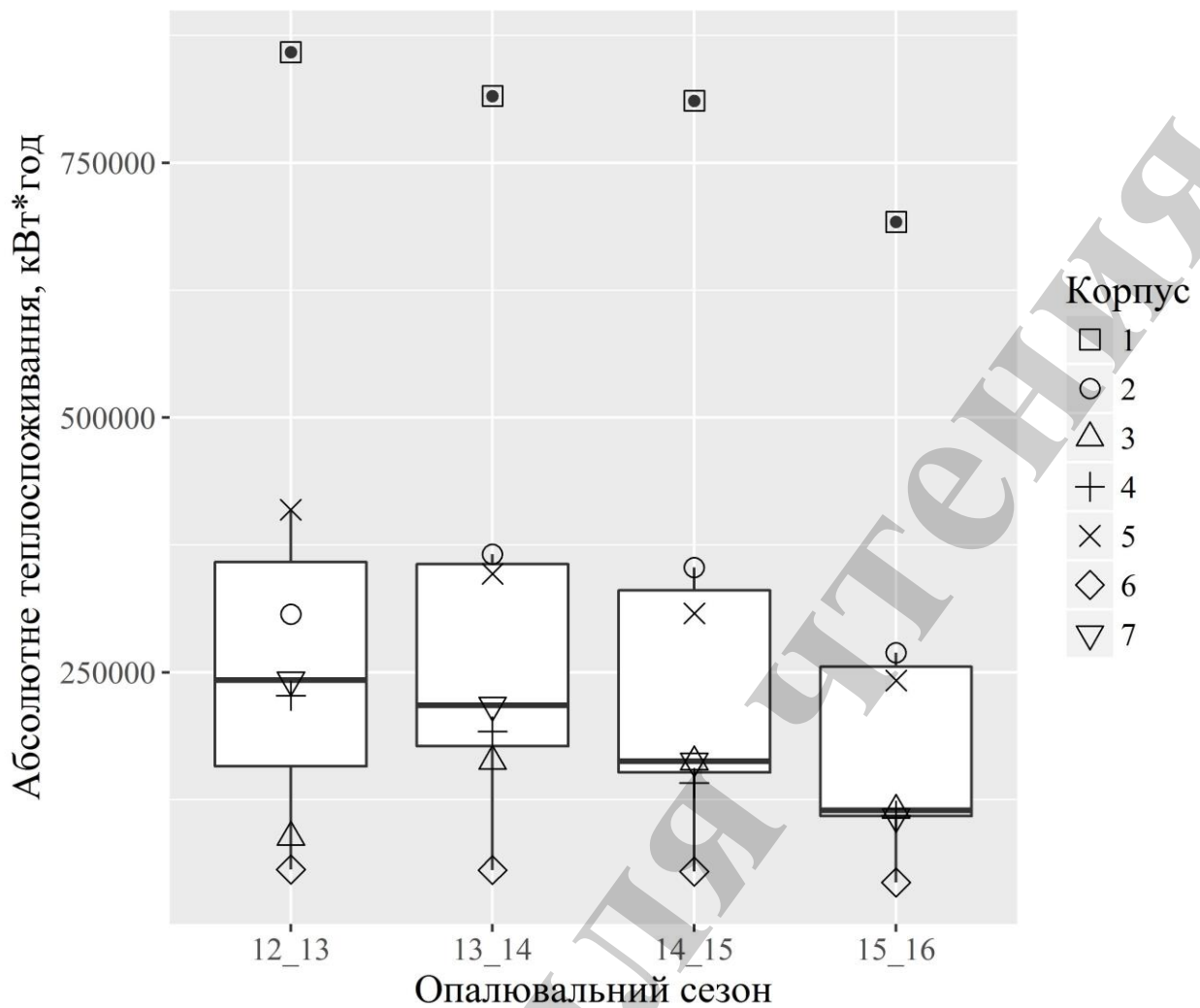


Рис. 10. Діаграма розмаху сумарних сезонних даних абсолютного теплоспоживання будівель

У наступних сезонах другий корпус перевищив показники п'ятого: в опалювальному сезоні 2014–2015 абсолютне теплоспоживання другого корпусу склало 352956 кВт·год (дозволило вийти за межі третього квартиля – 330374 кВт·год).

Питоме теплоспоживання третього корпусу в опалювальному сезоні 2012–2013 мінімальне і є викидом, хоча аналіз абсолютного теплоспоживання не дозволяє це визначити.

Питоме теплоспоживання четвертого корпусу за опалювальний сезон 2012–2013 більше цієї ж величини другого корпусу, враховуючи те, що теплове навантаження другого корпусу майже вдвічі перевищує теплове навантаження четвертого (рис. 11).

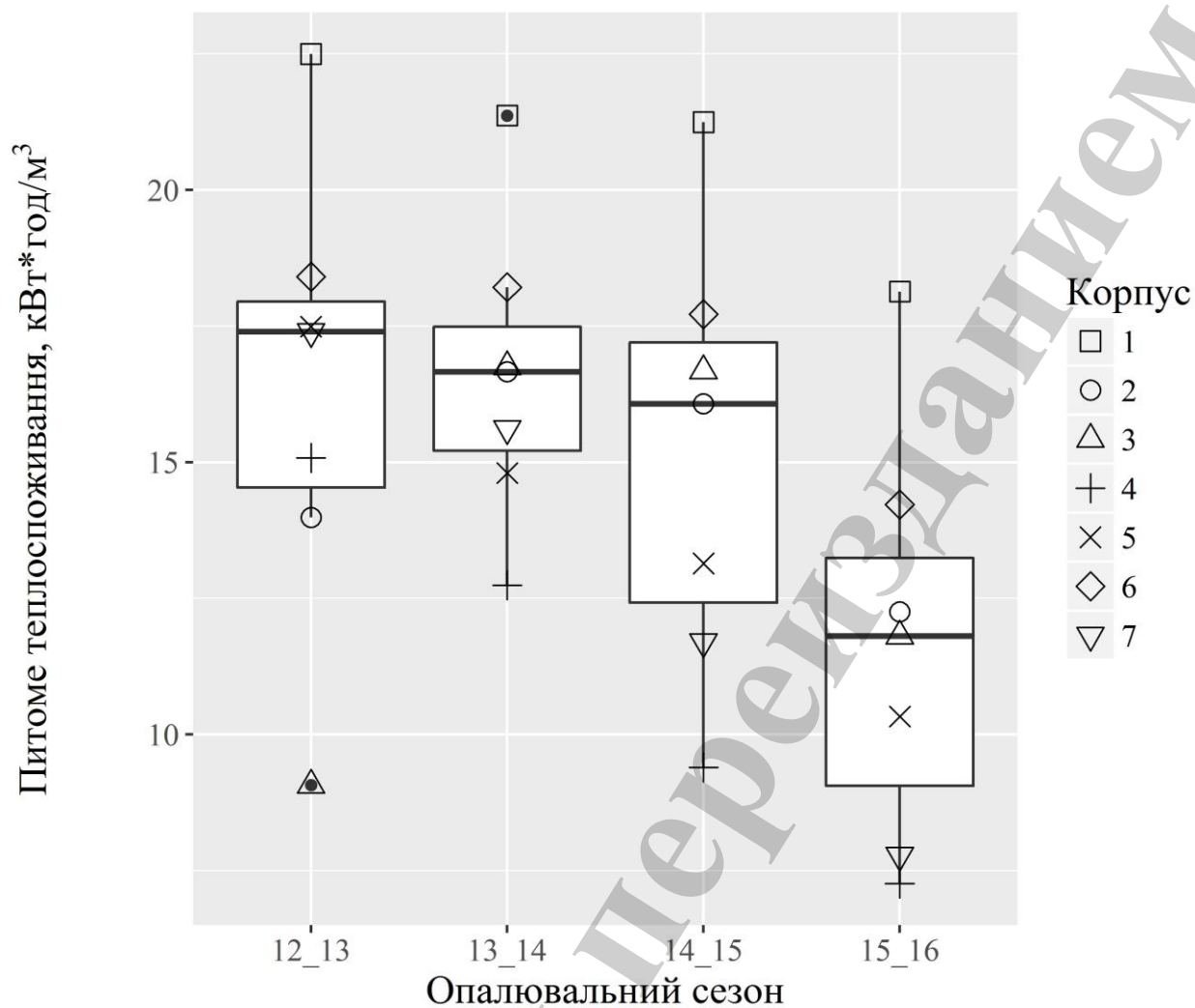


Рис. 11. Діаграма розмаху сумарних сезонних даних питомого теплоспоживання будівель

Облік питомого теплоспоживання, величина якого приведена до нормативних внутрішньої та зовнішньої температури повітря (рис. 12), дає можливість визначити зміни в тренді. В опалювальному сезоні 2013–2014 при максимальній середній температурі опалювального сезону (мінімальній кількості градусо-днів) відбувається істотне збільшення питомого теплоспоживання корпусів (викид по першому корпусу).

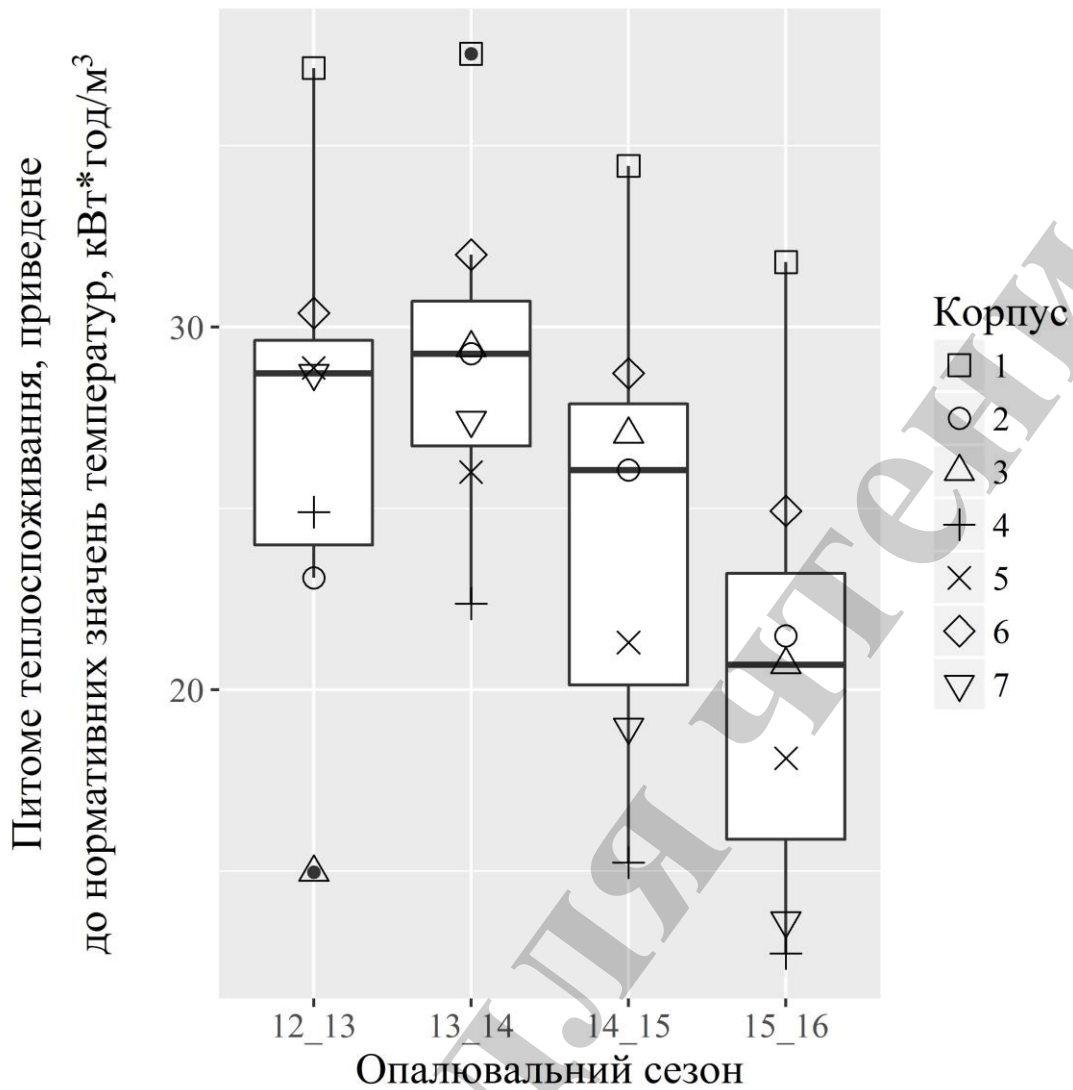


Рис. 12. Діаграма розмаху сумарних сезонних даних питомого теплоспоживання будівель, що приведені до нормативних значень зовнішньої та внутрішньої температури повітря

Питоме теплоспоживання за опалювальні сезони 2013–2014 і 2015–2016 з достатньо близькою кількістю градусо-днів значно відрізняється.

В опалювальному сезоні 2015–2016 відбувається значне зменшення абсолютного теплоспоживання у порівнянні з сезоном 2013–2014, враховуючи незначну різницю середньої температури повітря за опалювальні сезони.

Проведений аналіз абсолютного сумарного енергоспоживання за опалювальні періоди з 2012–2016 дає можливість говорити про тенденції до зниження споживання енергоресурсів корпусами. Незважаючи на це, значення абсолютного сумарного енергоспоживання першого корпусу є викидом (рис. 13).

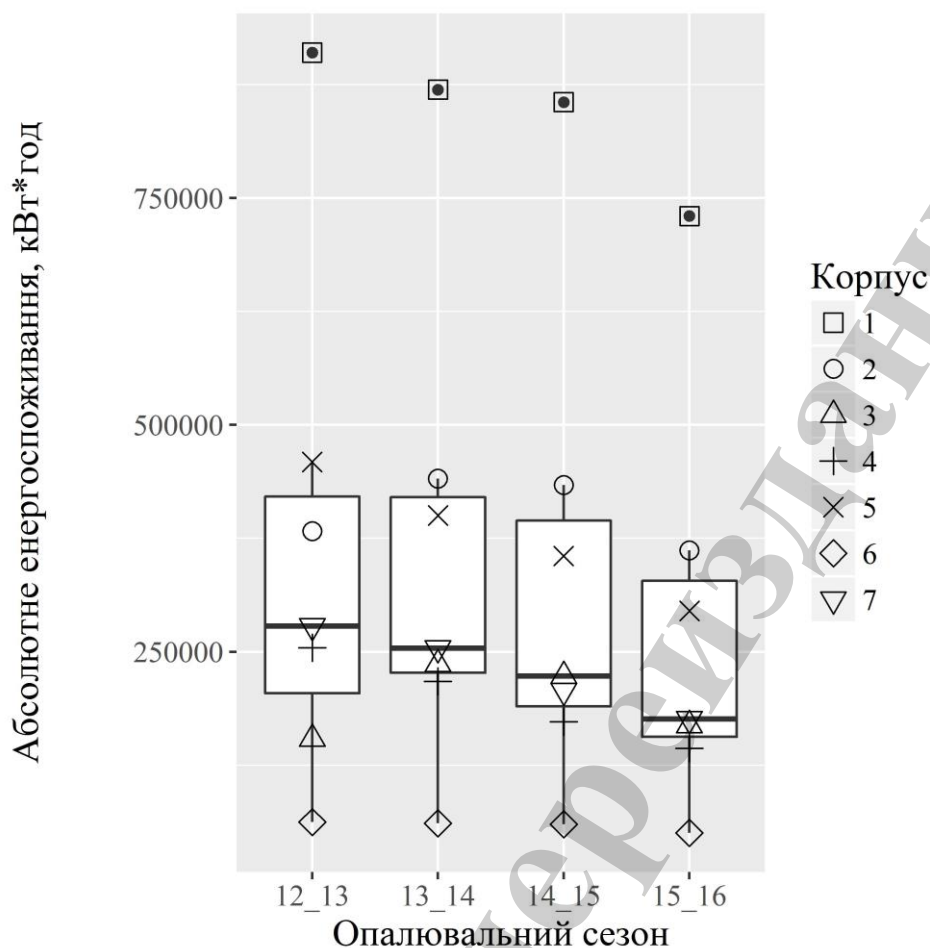


Рис. 13. Діаграма розмаху сумарних сезонних даних абсолютного енергоспоживання будівель

В опалювальному сезоні 2012–2013 споживання п'ятого корпусу перевищило споживання другого, з сезону 2013–2014 енергоспоживання другого корпусу більше п'ятого.

При переході до питомого сумарного енергоспоживання тренд зберігається, а залежність від температури на сезонних даних простежується незначно.

Спостерігаємо значний стрибок значення питомого енергоспоживання у третього корпусу, який на даних абсолютного сумарного енергоспоживання простежується слабкіше (у період з 2012–2014) (рис. 14).

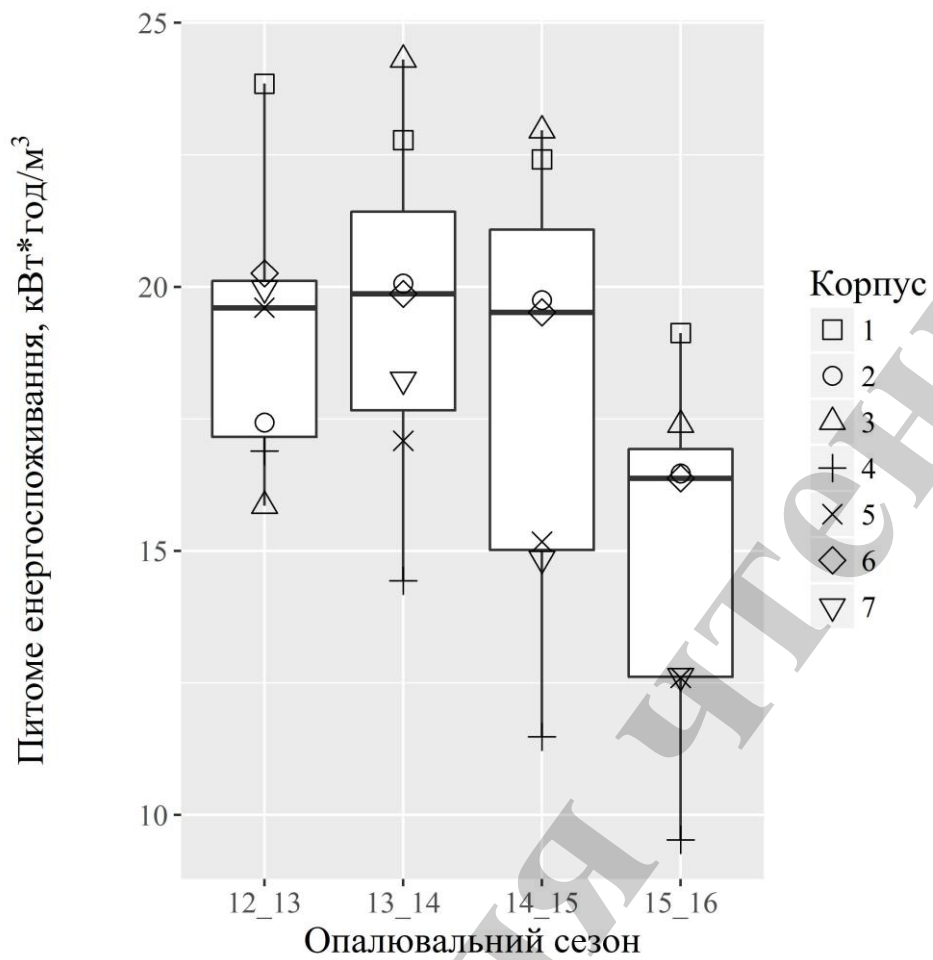


Рис. 14. Діаграма розмаху сумарних сезонних даних питомого енергоспоживання будівель

В опалювальних сезонах 2013–2014 і 2014–2015 споживання кВт·год на одиницю об'єму у третього корпусу перевищило цей показник першого корпусу, який має найбільший об'єм (рис. 15).

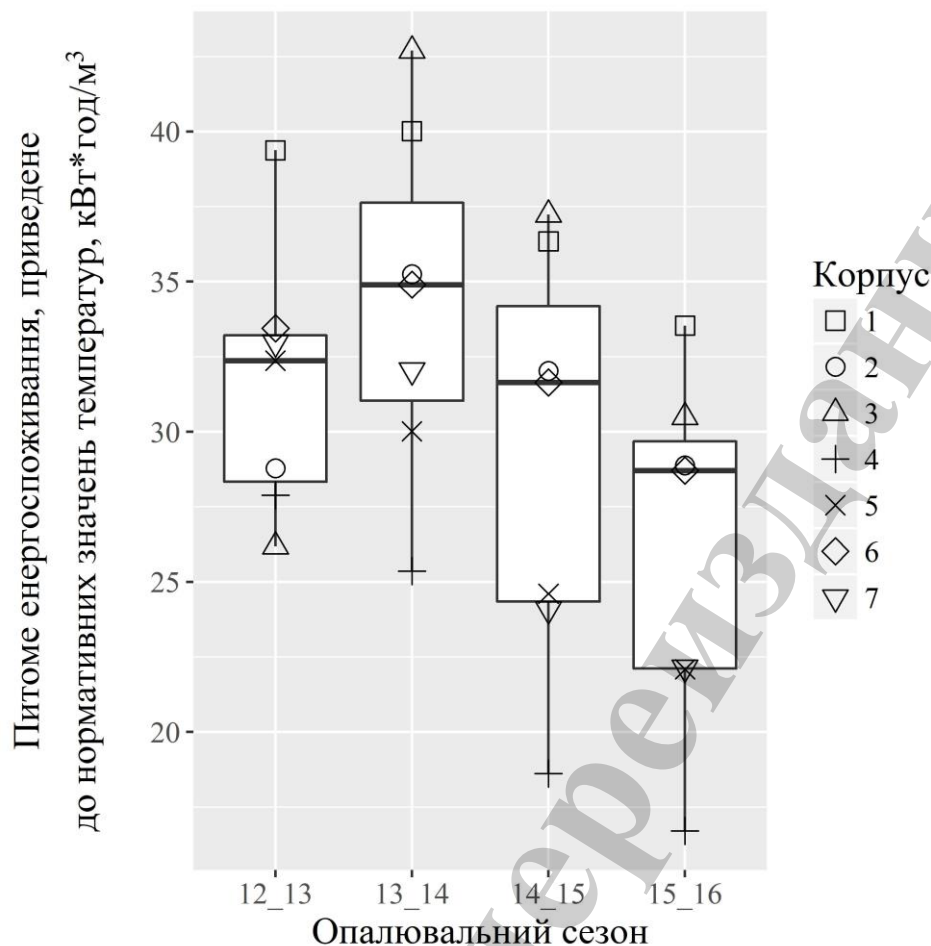


Рис. 15. Діаграма розмаху сумарних сезонних даних питомого енергоспоживання будівель, що приведені до нормативних значень зовнішньої та внутрішньої температури повітря

5. 3 Розробка web-базованого рішення для аналізу теплоспоживання квартир

Для спрощення процедури аналізу даних щодо теплоспоживання квартир розроблено відповідний web-додаток. У якості інструменту реалізації клієнт-серверної архітектури для дослідження даних щодо теплоспоживання пропонується фреймворк Shiny, який дозволяє швидко і зручну розробку веб-інтерфейсу із застосуванням open-source бібліотек і мови R.

Користувач з будь-якого пристрою, що має доступ до мережі Internet, може, виконавши процедуру авторизації або реєстрації, отримати можливість та інструментарій для обробки даних щодо теплоспоживання квартир. Дані, що вводяться користувачем, зберігаються в базі даних MySQL, причому, зважаючи на існування декількох типів користувачів, обмежується набір функцій, що може виконувати користувач. При спробі виконати функцію, що не входить до списку передбачених для даного користувача, виводиться повідомлення про помилку.

В залежності від типу будівлі, дані про теплоспоживання якої будуть досліджуватися, на основі створених структур інформаційних потоків, передбачаються різні форми даних.

Реалізована можливість завантаження користувачем файлу з місячними даними або введення у розроблену електронну форму, їх перерахунку в значення сезонного споживання та створення додаткових параметрів для порівняння: питомих споживання, коефіцієнт енергоефективності, тощо. Додатково зберігаються та виводяться на додатковий графік значення зовнішньої температури повітря, що отримані з Internet-ресурсів. Параметри для налаштування візуалізації дозволяють не тільки вибирати декартові вісі, а й здійснювати додаткове розділення по якісним змінним: кількість кімнат, наявність лічильника, тощо.

На рис. 16 представлено фрагмент web-інтерфейсу створеної програми з відображенням сезонних показників.

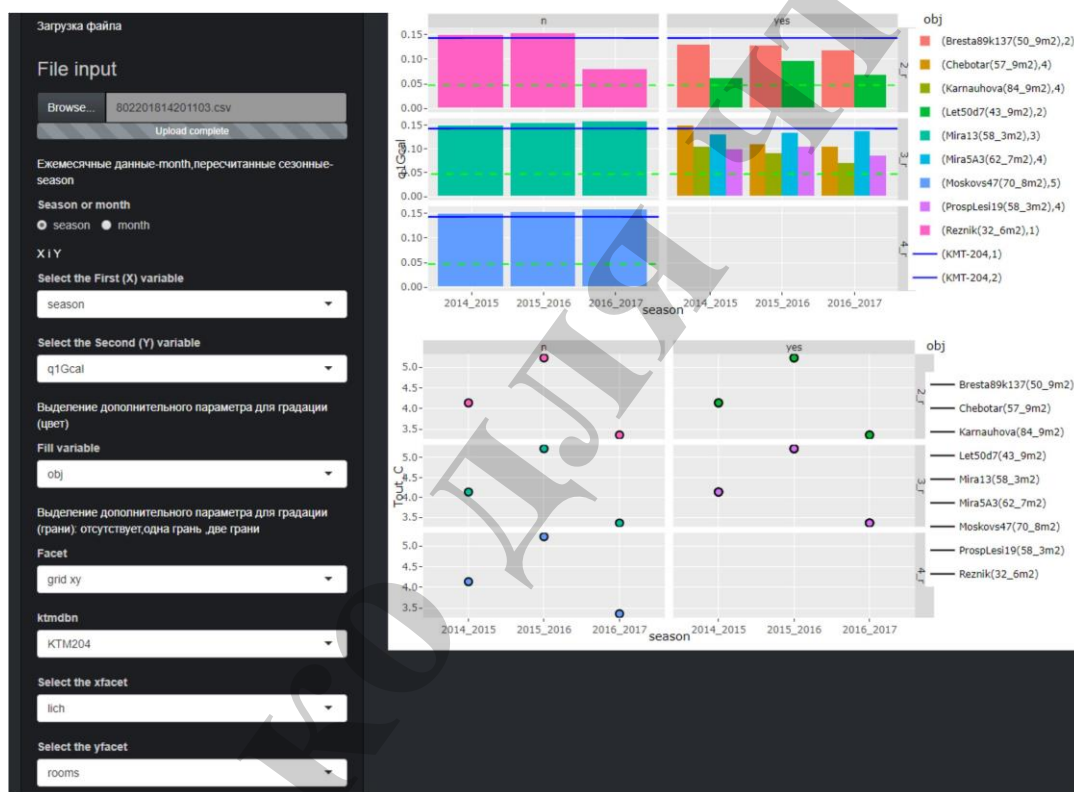


Рис. 16 Фрагмент web-інтерфейсу створеної програми з відображенням сезонних показників

Передбачена можливість додаткового виведення норм споживання теплової енергії для централізованого опалення (теплопостачання). Вирішення задачі створення та можливості збереження згенерованих звітів обумовлює необхідність забезпечення типографської якості в заданому форматі. Так ці звіти будуть зрозумілими користувачу, який не обов'язково є спеціалістом з енергоменеджменту. Реалізація базується на використанні мови розмітки даних – Markdown.

6. Обговорення результатів дослідження енергоспоживання будівель

Порівняння матриць кореляцій загальних даних (рис. 5), помісячних даних опалювальних (рис. 6) і неопалювальних сезонів (рис. 7) та сумарних даних опалювальних (рис. 8) і неопалювальних сезонів (рис. 9) дозволяє дійти висновків:

1. Перехід від загальних даних до даних опалювального періоду проявляє зв'язок між питомим теплоспоживанням та сумарним питомим енергоспоживанням, які приведені до нормативних внутрішньої та зовнішньої температур повітря.

2. Взаємозв'язок між абсолютними тепло- та електроспоживанням виражений слабко і зменшується при переході від загальних даних до даних опалювальних періодів.

3. Існує зв'язок між абсолютним та питомим теплоспоживанням. Додаткове врахування нормативних параметрів по температурі дещо погіршує цей зв'язок, проте додатково враховує об'єм та зовнішню температуру.

4. Абсолютне теплоспоживання достатньо залежить від об'єму будівель, як на загальних даних, так і на даних опалювальних періодів.

5. При переході до сумарних даних за опалювальні періоди зв'язок між питомими тепло- та енергоспоживанням стає виразнішим.

Проведений аналіз діаграм розмаху сумарних сезонних даних (рис. 10–15) вказує на тенденцію до зниження енергоспоживання, що викликана використанням автоматизованої системи моніторингу та керування теплоспоживанням у окремих будівлях. Спостерігається значне теплоспоживання адміністративним корпусом та перевитрата електроенергії у корпусі, що розміщує в собі їдальню.

Отримані результати початкового аналізу можуть слугувати основою для створення моделі, що, враховуючи структуру інформативних даних та цілі, зможе розв'язувати поставлені аналітичні задачі, візуалізувати результати та генерувати звіти.

Такі висновки можуть вважатися за доцільні з практичної точки зору, тому, що дозволяють обґрунтовано підходити до процесів енергомоніторингу та енергоменеджменту.

Необхідно відмітити, що в статті не наведений алгоритм роботи з пропущеними даними, не наведені обмеження на розмір вибірки, що продиктовані особливостями використання тесту на перевірку нормальності розподілу: кількість значень в діапазоні від 3 до 5000. До того ж існують певні обмеження на використання даної методики в початковому її вигляді, адже частина апріорних даних може бути недоступною. Так, дані про обсяг теплового навантаження отримуються з технічної документації на об'єкт, що може ускладнювати або унеможлиблювати доступ до неї, якщо йде мова про енергоспоживання не будівель в цілому, а окремих споживачів. Оскільки передбачена можливість генерації та зберігання типових звітів щодо початкового аналізу даних, які створюються з використанням технології грамотного програмування, створюється ряд суворих вимог до даних, що досліджуються. Конструювання нових ознак для порівняння енергоспоживання будівель спричиняє зростання розмірності да-

них, а отже, приріст часу виконання різного роду дій над даними, що може бути критичним для більших вибірок за значні періоди часу. Неможливість зняти названі обмеження в рамках даного дослідження породжує потенційно цікавий напрям подальших досліджень.

6. Висновки

Для підвищення ефективності теплопостачання необхідно розробити інформаційні рішення, спрямовані на інтелектуальну обробку даних щодо енергоспоживання та вирішення задач зменшення витрати енергоресурсів.

1. У ході дослідження розроблено структуру інформаційних потоків для отримання даних з доступних джерел. Встановлено особливості даних в залежності від типу та функціональних особливостей будівель. Дослідження структури даних підтвердило необхідність створення додаткових ознак для аналізу енергоспоживання будівель у вигляді питомого теплоспоживання з урахуванням температур зовнішнього та внутрішнього повітря, класу енергоефективності та тарифів на одиницю опалювальної площі.

2. З урахуванням складної структури даних щодо енергоспоживання будівель розроблено методику початкового аналізу на основі міжгалузевої методології CRISP-DM. Проведена апробація методики на даних навчальних будівель КрНУ дозволяє стверджувати, що даний алгоритм може використовуватися для попереднього аналізу даних на основі інструментарію Data Mining у подальшому.

3. Для поліпшення та полегшення аналізу даних теплоспоживання квартир було створено web-додаток. Із його допомогою здійснюється збір даних та візуалізація. Передбачена можливість автоматичного перерахунку існуючих помісячних даних у сезонні, створення нових параметрів для аналізу, завантаження календаря температури в залежності від географічного положення користувача тощо. Також можна порівнювати теплоспоживання квартир із встановленими вимогами до ефективного використання енергоресурсів за показником максимальної питомої потреби на опалення.

Література

1. A review of data-driven approaches for prediction and classification of building energy consumption / Wei Y., Zhang X., Shi Y., Xia L., Pan S., Wu J. et. al. // *Renewable and Sustainable Energy Reviews*. 2018. Vol. 82. P. 1027–1047. doi: <https://doi.org/10.1016/j.rser.2017.09.108>

2. Перекрест А. Л., Романенко С. С. Науково-прикладні аспекти енергоспокозбереження в комунальній енергетиці // *Електромеханічні і енергозберігаючі системи*. 2015. Вип. 2. С. 162–170.

3. Complex information and technical solutions for energy management of municipal energetics / Perekrest A., Shendryk V., Pijarski P., Parfenenko Y., Shendryk S. // *Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2017*. 2017. doi: <https://doi.org/10.1117/12.2280962>

4. Balco P., Drahošová M., Kubičko P. Data analysis in process of energetics resource optimization // *Procedia Computer Science*. 2018. Vol. 130. P. 597–602. doi: <https://doi.org/10.1016/j.procs.2018.04.109>

5. Zakovorotnyi A., Seerig A. Building energy data analysis by clustering measured daily profiles // *Energy Procedia*. 2017. Vol. 122. P. 583–588. doi: <https://doi.org/10.1016/j.egypro.2017.07.353>
6. Arregi B., Garay R. Regression analysis of the energy consumption of tertiary buildings // *Energy Procedia*. 2017. Vol. 122. P. 9–14. doi: <https://doi.org/10.1016/j.egypro.2017.07.290>
7. Energy consumption forecasting based on Elman neural networks with evolutive optimization / Ruiz L. G. B., Rueda R., Cuéllar M. P., Pegalajar M. C. // *Expert Systems with Applications*. 2018. Vol. 92. P. 380–389. doi: <https://doi.org/10.1016/j.eswa.2017.09.059>
8. Ahmad M. W., Mourshed M., Rezgui Y. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption // *Energy and Buildings*. 2017. Vol. 147. P. 77–89. doi: <https://doi.org/10.1016/j.enbuild.2017.04.038>
9. Jeffrey Kuo C.-F., Lin C.-H., Lee M.-H. Analyze the energy consumption characteristics and affecting factors of Taiwan's convenience stores-using the big data mining approach // *Energy and Buildings*. 2018. Vol. 168. P. 120–136. doi: <https://doi.org/10.1016/j.enbuild.2018.03.021>
10. Парфененко Ю. В., Шендрик В. В., Галіченко О. С. Прогнозування теплоспоживання будівель соціально-бюджетної сфери з використанням нейронних мереж // *Радіоелектроніка, інформатика, управління*. 2015. № 2. С. 41–46. URL: http://nbuv.gov.ua/UJRN/riu_2015_2_7
11. Slabchenko O., Sydorenko V., Siebert X. Development of models for imputation of data from social networks on the basis of an extended matrix of attributes // *Eastern-European Journal of Enterprise Technologies*. 2016. Vol. 4, Issue 2 (82). P. 24–34. doi: <https://doi.org/10.15587/1729-4061.2016.74871>
12. Building energy performance diagnosis using energy bills and weather data / Geng Y., Ji W., Lin B., Hong J., Zhu Y. // *Energy and Buildings*. 2018. Vol. 172. P. 181–191. doi: <https://doi.org/10.1016/j.enbuild.2018.04.047>
13. A study and a directory of energy consumption data sets of buildings / Babaei T., Abdi H., Lim C. P., Nahavandi S. // *Energy and Buildings*. 2015. Vol. 94. P. 91–99. doi: <https://doi.org/10.1016/j.enbuild.2015.02.043>
14. Перекрест А. Л., Загирняк М. В. Опыт внедрения и использования автоматизированной системы мониторинга температурных режимов и удаленного управления теплоснабжением Кременчугского национального университета // *Электротехнические и компьютерные системы*. 2014. № 15 (91). С. 423–426.