

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
„КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО”**

**ФАКУЛЬТЕТ ЕЛЕКТРОНІКИ
КАФЕДРА ЕЛЕКТРОННИХ ПРИСТРОЇВ ТА СИСТЕМ**

До захисту допущено
Завідувач кафедри

(підпис) Юлія ЯМНЕНКО
(ім'я ПРІЗВИЩЕ)

" ____ " _____ 2021 р.

Дипломний проєкт

на здобуття першого (бакалаврського) рівня вищої освіти

Спеціальність 171 Електроніка
(код та назва спеціальності)

Освітня програма Електронні компоненти і системи

на тему: Прогнозування навантаження у Microgrid

Виконав: студент IV курсу, групи ДС – 71

Ростислав МАРЧЕНКО
(ім'я ПРІЗВИЩЕ) _____
(підпис)

Керівник доцент каф. ЕПС, к.т.н., Вадим ВОЛКІВСЬКИЙ
(посада, науковий ступень, вчене звання, ім'я ПРІЗВИЩЕ) _____
(підпис)

Консультант технічний розділ
доцент каф. ЕПС, к.т.н., доц. Катерина КЛЕН
(назва розділу) (посада, науковий ступень, вчене звання, ім'я ПРІЗВИЩЕ) _____
(підпис)

Рецензент завідувач каф. АМЕС, д.т.н., проф. Сергій НАЙДА
(посада, науковий ступень, вчене звання, ім'я ПРІЗВИЩЕ) _____
(підпис)

Консультант
з нормоконтролю доцент каф. ЕПС, к.т.н., доц. Лариса БАТРАК
(посада, науковий ступень, вчене звання, ім'я ПРІЗВИЩЕ) _____
(підпис)

Засвідчую, що у цьому дипломному
проєкті немає запозичень з праць інших
авторів без відповідних посилань

Студент _____
(підпис)

Київ – 2021 року

№ з/п	Формат	Позначення			Найменування	Кількість аркушів	Примітка		
1					Завдання на дипломний проєкт	1			
2	A4	ДС.7110.001 ПЗ			Пояснювальна записка	69			
3									
4	A1	ДП.ДС71.01.000 Е2			Схема електрична структурна	1			
5	A1	ДП.ДС71.01.000 ТК			Теоретичне креслення	1			
6	A1	ДП.ДС71.01.000ТК			Теоретичне креслення	1			
</									

**Національний технічний університет України
“Київський політехнічний інститут
імені Ігоря Сікорського”**

Факультет електроніки
(повна назва)

Кафедра електронних пристроїв та систем
(повна назва)

Рівень вищої освіти – перший (бакалаврський)

Спеціальність 171 Електроніка
(шифр і назва)

Спеціалізація Електронні компоненти і системи

ЗАТВЕРДЖУЮ

В.о. завідувача кафедри

(підпис) Юлія ЯМНЕНКО
(ім'я ПРІЗВИЩЕ)

"20" травня 2021 р.

**З А В Д А Н Н Я
НА ДИПЛОМНИЙ ПРОЄКТ СТУДЕНТУ**

Ростиславу МАРЧЕНКУ
(ім'я ПРІЗВИЩЕ)

1. Тема проєкту Прогнозування навантаження у Microgrid

Керівник проєкту доцент каф. ЕПС, к.т.н., Вадим ВОЛКІВСЬКИЙ,
(посада, науковий ступень, вчене звання, ім'я ПРІЗВИЩЕ)

затверджені наказом по університету від « 24 » травня 2021 року № 1316-с

2. Термін подання студентом проєкту 07.06.2021 року

3. Вихідні дані до проєкту Дані навантаження отримані для будинку; дані навколишнього середовища; дальність прогнозу – одна доба.

4. Зміст пояснювальної записки (перелік завдань, які потрібно розробити) 1) Методи прогнозування навантаження у microgrid. 2) Побудова прогнозної моделі. 3) Програмна реалізація прогнозування навантаження

5. Перелік графічного (ілюстративного) матеріалу (із зазначенням обов'язкових креслеників, плакатів, презентацій тощо) Схема електрична структурна; Математична модель прогнозування навантаження; Блок-схема алгоритму та результати роботи програми.

6. Консультанти розділів проєкту

Розділ	Ім'я ПРИЗВИЩЕ, посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Технічний	Катерина КЛЕН, доцент кафедри ЕПС	20.05.2021	07.06.2021

7. Дата видачі завдання 20.05.2021

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів виконання дипломного проєкту	Строки виконання етапів проєкту	Примітка
1	Огляд методів прогнозування	11.04-18.04	Виконано
2	Оцінка прогнозування навантаження	11.04-18.04	Виконано
3	Похибки прогнозування	18.04-25.04	Виконано
4	Кореляційні коефіцієнти	25.04-02.05	Виконано
5	Добові криві зміни потужності навантаження	02.05-09.05	Виконано
6	Оцінка подібності днів за даними навколишнього середовища	09.05-16.05	Виконано
7	Побудова прогнозової моделі за регресивними рівняннями	16.05-19.05	Виконано
8	Розробка блок-схеми алгоритму	19.05-23.05	Виконано
9	Написання коду програми	23.05-30.05	Виконано
10	Оформлення пояснювальної записки	30.05-07.06	Виконано

Студент

(підпис)

Ростислав МАРЧЕНКО

(ім'я ПРИЗВИЩЕ)

Керівник проєкту

(підпис)

Вадим ВОЛКІВСЬКИЙ

(ім'я ПРИЗВИЩЕ)

АНОТАЦІЯ

На поточний день виникає питання прогнозування навантаження локальних об'єктів. Через зростання таких систем як Microgrid, та прагнення суспільства до більш природних джерел живлення. Тому спосіб прогнозування навантаження з якомога меншими похибками залишається актуальним. У дипломній роботі були досліджені існуючі методи прогнозування навантаження локальних об'єктів. Та способи оцінки адекватності прогнозованої моделі.

Було обрано метод регресійного аналізу, оскільки при цьому можливо створити автоматизовану систему прогнозу. Крім того поведінка кривої енергоспоживання описується функцією часу, тому обираємо групу моделей часових проміжків.

Розроблено програмну реалізацію прогнозування навантаження в залежності від навколишнього середовища, похибка прогнозу становить 40%.

Ключові слова: прогнозування навантаження, регресійний аналіз, кореляція, Micro Grid, системи розподіленої генерації.

ANNOTATION

At present, there is a question of forecasting the load of local objects. Due to the growth of systems such as Microgrid, and society's desire for more natural power sources. Therefore, the method of forecasting the load with the smallest possible errors remains relevant. In the diploma work the existing methods of forecasting of loading of local objects were investigated. And ways to assess the adequacy of the predicted model.

The method of regression analysis was chosen because it is possible to create an automated forecasting system. In addition, the behavior of the energy consumption curve is described as a function of time, so we choose a group of models of time intervals.

A software implementation of load forecasting depending on the environment has been developed, the forecast error is 40%.

Keywords: load of forecasting, regression analysis, correlation, Micro Grid, distributed generation systems.

ЗМІСТ

ВСТУП	3
1. МЕТОДИ ПРОГНОЗУВАННЯ НАВАНТАЖЕННЯ У MICROGRID.....	4
1.1 Огляд методів прогнозування	4
1.2 Оцінка прогнозування навантаження	14
1.3 Похибки прогнозування	16
2. ПОБУДОВА ПРОГНОЗНОЇ МОДЕЛІ.....	20
2.1 Добові криві зміни потужності навантаження.....	20
2.2 Оцінка подібності днів за даними навколишнього середовища	28
2.3 Побудова прогнозної моделі за регресивними рівняннями.....	30
3. ПРОГРАМНА РЕАЛІЗАЦІЯ ПРОГНОЗУВАННЯ НАВАНТАЖЕННЯ	41
3.1 Опис алгоритму	41
3.2 Особливості мови програмування.....	45
3.3 Результати роботи програми.....	47
ВИСНОВКИ.....	51
ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ	52
SUMMARY	

					ДП.ДС71.10.000ПЗ				
Змн	Арк А	№ докум №	Підпис П	Дата					
Розроб	Марченко				Прогнозування навантаження у Microgrid. Пояснювальна записка	Літ	Літ	Арк	Арк вів
Перевір	Волківський							7	69
Реценз.						«КПІ ім. Ігоря Сікорського», ФЕЛ, ЕПС, гр. ДС-71			
Н. Контр.	Батрак Л.М								
Затверд.	Волківський								

ВСТУП

На поточний день виникає питання прогнозування навантаження локальних об'єктів, за даними навколишнього середовища. Це актуально для компаній, які займаються електропостачанням. Ці компанії зобов'язані звітувати плановий погодинний графік електроспоживання на добу вперед. Результатом відхилення від планового графіку це додаткові фінансові втрати на закупівлю, або ж продажу надлишкових об'ємів електроенергії по низькій ціні. Також, прогнозування електроспоживання є актуальним для промислових об'єктів і приватних будинків. На даний момент прогнозування найчастіше використовують в Microgrid - це локалізоване угруповання розподілених джерел енергії, таких як сонячна енергія, вітер, гідроелектростанції, а також інструменти накопичення та резервування та управління навантаженням. Ця система може працювати як окрема сутність, або її користувачі можуть підключатися до більшої мережі за необхідності. Зростання мікромереж очікується в розвинених країнах, що посилюють стійкість електричних мереж, а також в країн, що розвиваються в районах, які в даний час не мають доступу до електроенергії, використовуючи альтернативні джерела відновлюваної енергії, такі як гідроенергія, розподілена сонячна фотоелектрична енергія. Передбачається, що ці системи замінюють те, що в іншому випадку було б продовженням брудної електромережі або подальшим використанням поза мережевих масляних або дизельних генераторів.

Тому, прогнозування навантаження буде завжди актуальним особливо при розповсюдженні таких систем як Microgrid.

					ДП.ДС71.10.000ПЗ	Арк.
						3
Змн.	Арк.	№ докум.	Підпис	Дата		

1. МЕТОДИ ПРОГНОЗУВАННЯ НАВАНТАЖЕННЯ У MICROGRID

1.1 Огляд методів прогнозування

Метод прогнозування – це спосіб дослідження об'єкта з певною послідовністю дій, які необхідно виконати для отримання моделі прогнозування.

Методи прогнозування поділяються на інтуїтивні, формалізовані, комбіновані (об'єднують інтуїтивні і формалізовані методи) рис.1.1.

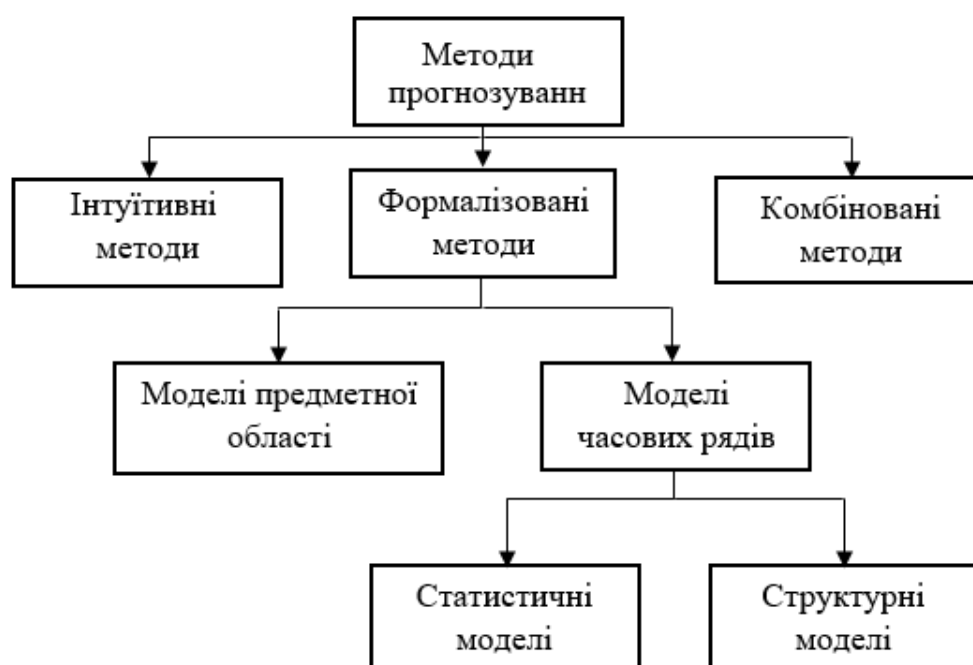


Рис.1.1. Класифікація методів прогнозування

Інтуїтивні методи ґрунтуються на інформації колективних експертних оцінок, цей метод корисний в процесі створення прогнозної моделі навантаження. Точність прогнозу залежить від кваліфікації експерта, але через високу ступінь залежності від якості прогнозу спеціаліста, можемо дійти висновків, що інтуїтивні методи не підходять для створення моделей прогнозу навантаження. Також інтуїтивні методи,

корисні при створенні прогностної моделі на основі формалізованих методів [1].

Формалізовані методи використовують при достатній кількості зібраної інформації, а саме про теперішній і минулий стан об'єкту дослідження.

Формалізовані методи поділяються на :

- математичні методи (регресивний аналіз, факторний аналіз);
- методи асоціативності (методи імітаційного моделювання і архайчного аналізу);
- методи екстраполяції (методи найменших квадратів, методи експоненціального згладжування, методи адаптивного згладжування);
- структурно-системні методи (методи функціонально-ієрархічного моделювання, методи морфологічного аналізу, матричний метод).

Також, одним з різновидів формалізованих методів є моделі часових рядів, які шукають залежності всередині самого процесу. Моделі часових рядів застосовують для різних предметних областей : механіки, термодинаміки, електротехніки [2].

Комбіновані методи – це методи зі змішаною інформаційною основою, в яких в якості первинної використовується як експертна, так і фактографічна інформація [3]. Наприклад, експерти висувають колективні сценарії, потім розглядають певну кількість тенденцій розвитку можливих подій. У сценарії робиться спроба проаналізувати без кількісних оцінок мети підприємства спрямування зусиль і завдання. Отримані висновки використовуються для створення "дерева цілей" та встановлення коефіцієнтів значущості [3]. "Дерева цілей" дозволяють знаходженню додаткових характеристик, таких як: коефіцієнтів використання отриманих результатів в інших областях .

Розглянувши представлені методи, можемо дійти висновку що для прогнозування навантаження у MicroGrid найбільш підходящим є

					ДП.ДС71.10.000ПЗ	Арк.
						5
Змн.	Арк.	№ докум.	Підпис	Дата		

фактографічні (формалізовані) методи, також навантаження описується функцією часу, тому підходять моделі часових рядів.

Прогнозування часових рядів

Часовий ряд - це набір даних з певною послідовністю в різні моменти часу [4].

Прогнозування часових рядів — це застосування моделі для передбачування майбутніх значень на основі значень попередньо спостережених. Часові ряди поділяють на статистичні та структурні.

Статистичний метод допомагає знайти залежність між навантаженням і зовнішнім фактором який вплинув на дані в минулому і завдяки чому можемо отримати рівняння залежності між навантаженням та чинником впливу. Тому статистичні методи можуть передбачати навантаження будь якого дня.

Найбільш поширені методи прогнозування часових рядів:

- кореляційний і регресійний аналіз;
- прогнозна екстраполяція;
- адаптивні методи прогнозування;
- авторегресивний аналіз;
- експоненціальне згладжування;
- класифікаційно-регресійні дерева;

Кореляційний і регресійний аналіз - це поширений метод прогнозування який базується на методі найменших квадратів.

Модель множинної регресії для лінійного випадку

$$y_j = \sum_i^n a_i x_{ij} + \varepsilon_j, \quad (1.1)$$

де a_i коефіцієнт регресії; y_j значення функції незалежної змінної, x_{ij} значення функції залежної зміни; n – кількість незалежних змінних ; ε_j - випадкова похибка [6].

					ДП.ДС71.10.000ПЗ	Арк.
						6
Змн.	Арк.	№ докум.	Підпис	Дата		

Для побудови моделі множинної регресії необхідно визначити коефіцієнти регресії, використовуючи відомості про об'єкт дослідження і чинники які на нього впливають. З цього випливає що кореляційні і регресійні методи здатні враховувати різноманітні параметри які впливають на об'єкт прогнозу .

Множинна регресійна модель. На $Z(t)$ впливає ряд дискретних зовнішніх факторів $X_1(t), \dots, X_n(t)$. Отже модель прогнозування має вигляд

$$Z(x) = \infty_0 + \infty_1 X_1(t) + \infty_2 X_2(t) + \dots + \infty_n X_n(t) + \varepsilon_t, \quad (1.2)$$

Щоб обчислити майбутнє значення процесу $Z(t)$ потрібно мати майбутні значення всіх вхідних факторів моделі. Якщо використовувати лінійні регресивні моделі то результат прогнозу буде отриманий швидше в порівнянні з іншими моделями, також доступність аналізу всіх можливих обчислень. Але є складності у визначенні виду функціональної залежності і параметрів моделі [5].

Прогнозна екстраполяція - це тимчасовий ряд представлений у вигляді

$$y_j = f(x) + \varepsilon_j, \quad (1.3)$$

де ε_j - випадкова величина а з нульовою середньою і дисперсією $D(\varepsilon_j)$; $f(x)$ - деяка не випадкова функція час.

Складний запис функції тимчасового ряду

$$y_j = f(x) + S + C + \varepsilon_j, \quad (1.4)$$

де $f(x)$ - компонента характеризує розвиток процесу; S - сезонна складова; C - циклічна складова; ε_j - компонента характеризує випадкові коливання процесу, що прогнозується. Тому щоб розрахувати модель прогнозування часового ряду необхідно визначити детерміновані компоненти ,стохастичні складові і сезонно-циклічні складові.

Переваги методу: простота реалізації, реалізація прогнозування з мінімальним набором вхідних даних.

					ДП.ДС71.10.000ПЗ	Арк.
						7
Змн.	Арк.	№ докум.	Підпис	Дата		

Недоліки: жорстка фіксація тренду моделі, яка не дозволяє врахувати короткострокові зміни навантаження [7].

Автогресивні моделі є найбільш популярними з статистичних моделей [8]. Це метод в якому часовий ряд знаходиться в лінійній залежності від якоїсь умови. Особливість цього методу в тому, що моделі беруть за основу минулі данні прогнозованих рядів, що зменшує варіативність прогнозу. Приклади моделей (ARIMA, ARIMAX, GARCH, ARDLM) для прогнозування використовуючи часові ряди різних предметних областей[9].

Переваги методу: отримання моделі з адекватним прогнозом при не великих вимог до вихідними даними а також при малій витраті часу .

Недоліки методу: прогноз за вихідними значеннями створюється лише на один період вперед. Якщо необхідно зробити прогноз на більший період часу, доведеться використовувати не реальні фактори, які мають вплив на розрахунки, а ті що отримали в результаті спрогнозуванні моделі. В результаті отримаємо прогноз на прогнозі, з цього впливає зменшення адекватності прогнозу. Для збільшення розрядності авторегресії необхідно розширити діапазон вихідних даних.

Адаптивні методи – це методи які враховують динаміку зміни характеристик процесів, що змінюються і в результаті пристосовуються до цієї зміни. Завдяки пристосуванню до динамічних змін, цей метод використовують в різних сферах діяльності . Сам прогноз створюється після збору даних спостереження об'єкта прогнозування. Отримавши достатню кількість даних, робиться прогноз, результат якого буде порівнюватись з фактичним результатом спостережень. Після порівняння створюється коригувальна модель щоб зробити прогноз для наступного спостереження і так до кінця всіх спостережень. Це дозволяє моделі пристосовуватися до нової інформації, і за певний період досліджень відображає тенденції об'єкта прогнозування.

					ДП.ДС71.10.000ПЗ	Арк.
						8
Змн.	Арк.	№ докум.	Підпис	Дата		

Головною перевагою адаптивних методів є здатність розрахувати точний прогноз на проміжок більший, ніж при використанні методу експоненціального згладжування. Якщо модель має довгий тимчасовий ряд, це дає можливість для застосування адаптивного методу, щоб отримати довгострокове прогнозування [9].

Також для прогнозування часових рядів, застосовують [10]:

- Моделі однофакторного прогнозування (моделі прогнозування тренду, методи прогнозування випадкової компоненти).
- Моделі багатфакторного прогнозування (адаптивні регресії, факторний аналіз, багатфакторна фільтрація).

Модель експоненціального згладжування застосовується для моделювання різноманітних процесів [15].

Основна ідея експоненціального згладжування полягає в постійній перевірці прогнозних даних при їх надходженні. Ця модель фіксує спостереження що починають експоненціально спадати. Тому, останні фіксовані спостереження мають більший вплив на прогнозне значення, в порівнянні з попередніми.

Функція моделі експоненціального згладжування має вигляд (1.5):

$$Z(t) = (a \cdot Z(t-1) + (t-a) \cdot S(t-1)) + \varepsilon_j, \quad (1.5)$$

де a - згладжуваний коефіцієнт. Кожне наступне згладжене значення $S(t)$ є середнім між попереднім значенням тимчасового ряду $Z(t)$ і згладженого значення $S(t-1)$.

Подвійне експоненціальне згладжування – ця модель використовується при моделюванні процесів, які мають тренд. В цьому випадку в моделі необхідно розглядати рівень і тренд [16].

Рівень і тренд згладжуються окремо:

$$Z(t) = S(t) + \varepsilon_j.$$

$$S(t) = a \cdot Z(t-1) + (t-a) \cdot (S(t-1) - B(t-1)).$$

					ДП.ДС71.10.000ПЗ	Арк.
						9
Змн.	Арк.	№ докум.	Підпис	Дата		

$$B(t) = \gamma \cdot (S(t-1) - S(t-2) + (1 + \gamma) \cdot B(t-1)), \quad (1.6)$$

де a - згладжуваний коефіцієнт, γ – коефіцієнт згладжування тренда.

Моделі на базі ланцюгів Маркова розглядають майбутній стан процесу, що залежить тільки від його поточного стану і одночасно не залежить від попередніх [15]. Саме через цю особливість ланцюгами Маркова, відносять до процесів з короткою пам'яттю. Ланцюг Маркова для процесу, з трьома станами, представлений на рис. 1.2 [17].

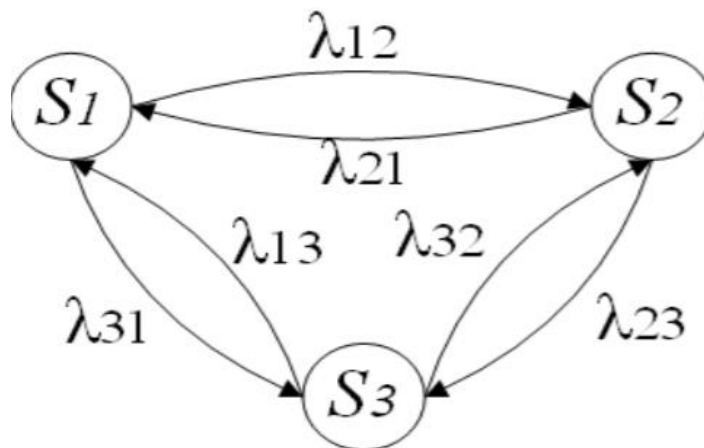


Рис. 1.2 Ланцюг Маркова з трьома станами

Тут $S_1, \dots, S_2$ - стану процесу $Z(t)$; λ_{23} – це ймовірність переходу з одного стану S_1 в інший стан S_2 , λ_{23} – це ймовірність переходу з одного стану S_2 в інший стан S_3 . Побудова ланцюга Маркова визначає безліч станів і ймовірності переходів [17].

Таким чином, ланцюг Маркова та ймовірність переходу станів характеризуються залежність між майбутнім значенням і поточним значенням процесу.

Моделі на базі регресійних дерев

Класифікаційне регресійне дерево – це структурна модель прогнозування часових рядів [18].

CART моделі використовують, щоб моделювати процеси, які знаходяться під впливом безперервних зовнішніх факторів та категоріальні. Якщо необхідно враховувати фактори двох типів, то використовуються змішані класифікаційні - регресійні дерева [19].

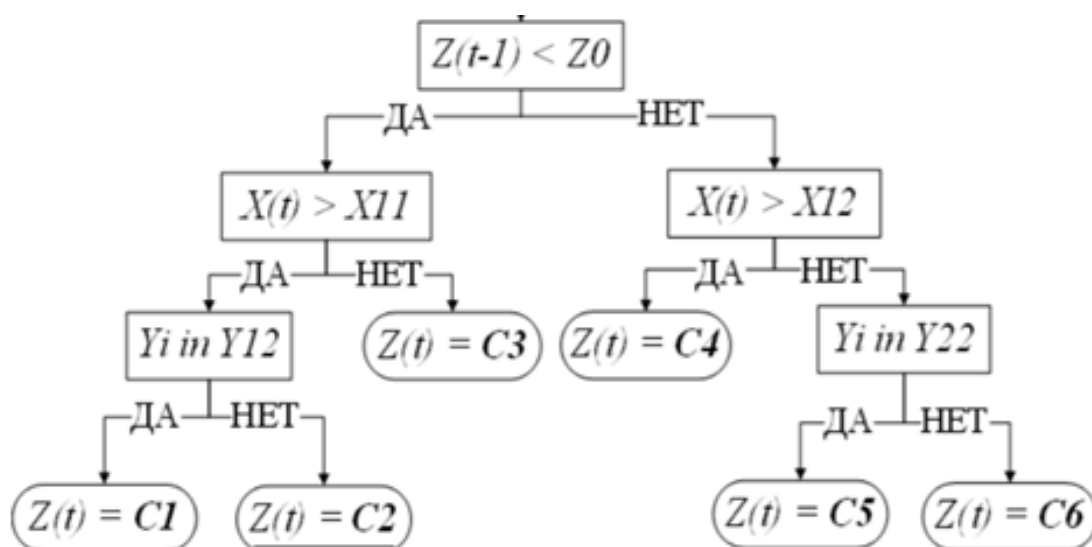


Рис.1.3- Бінарне класифікаційне - регресійне дерево

Згідно з CART моделлю, результат прогнозу часового ряду має залежність від деяких незалежних змінних і попередніх значень.

Розглянемо бінарне класифікаційне - регресійне дерево що наведено на рис.1.3, перше значення процесу порівнюється з Z_0 (константа). Коли $Z(t-1) < Z_0$, то відбувається наступна перевірка:.. Якщо в цьому випадку нерівність не виконується, переходимо до $Z(t) = C_3$. Таким чином перевірки продовжуються, поки не буде знайдений лист дерева, в якому визначається майбутнього значення процесу $Z(t)$.

Головне, що при визначенні значення приймаються безперервні змінні $X(t)$, так і категоріальні Y , які перевіряють присутність значення в розрахованих підмножин виконується перевірка присутності значення в одному з попередньо визначених підмножин. Значення порогових констант, Z_0 , X_{11} , підмножин Y_{11} , Y_{12} виконується на етапі навчання

дерева [18]. Отже, модель CART використовує: структури дерева, а також порогові константи та підмножини, для моделювання залежності майбутніх процесів $Z(t)$.

Переваги і недоліки моделей прогнозування

Розглянувши основні методи прогнозування у MicroGrid, можемо зробити висновок, що найбільш популярні статистичні методи такі як регресія і часові методи. Порівняння моделей прогнозування табл.1.1 [3].

Таблиця 1.1

Порівняння моделей прогнозування

Моделі	Переваги	Недоліки
Регресивні	Простота, гнучкість, прозорість моделювання; одноманітність аналізу і проектування	Складність визначення функціональної залежності; трудомісткість знаходження коефіцієнтів залежності;
Авторегресійні	Простота, прозорість моделювання нелінійних процесів моделювання; одноманітність аналізу і проектування; безліч прикладів застосування	Велика трудомісткість і ресурсомісткість ідентифікації; неможливість моделювання нелінійностей; низька адаптивність
Моделі експоненційного згладжування	Простота моделювання; одноманітність аналізу і проектування	Недостатня гнучкість; вузька застосовність моделей
Прогнозна екстраполяція	простота реалізації, реалізація прогнозування з мінімальним набором вхідних даних	жорстка фіксація моделі тренда, яка не дозволяє врахувати короткострокові зміни навантаження

Розглянувши існуючі сучасні методи прогнозування навантажень в розподільних мережах, робимо висновок, про те, що одного універсального методу, який відповідає всім вимогам не існує.

Кожен метод має свої переваги, недоліки, межі застосування.

До обмежень аналітичних методів відносять:

- трудомісткість, складність обчислень, які необхідно обирати та обчислювати окремо для параметрів аналізованого виразу;

- не точність результатів прогнозування при не правильному виборі моделі.

Для коректного використання регресійного аналізу слід дотримуватися певних умов:

- факторні ознаки повинні задані точно і корельовано. Результуюча ознака повинна мати постійну дисперсію;
- шукаючи залежність між погодними змінними і навантаженням, отримаємо нелінійні залежності.

До авторегресивних моделей часових рядів є вимогою виконання статистичних критеріїв:

- критерій стаціонарності і випадковості часового ряду,
- критерій значущості кореляційних коефіцієнтів.

Метод експоненціального згладжування застосовується, якщо під час моделювання в результаті якого процес має монотонний характер, а сам часовий ряд процесу має характеристику тимчасового короткострокового ряду з числом значень до 100.

Недоліки :

- неможливість прямого обліку в моделі дії конкретного впливового фактору;
- необхідність використання прогнозуючих поліномів порядку першого, другого порядку, через ускладнення формул ідентифікації та відсутності аналітичного виведення подібних формул.

Моделі і методи на базі ланцюгів Маркова

Перевагою цього методу є простота і одноманітність аналізу та проектування. Головний недолік цього методу – відсутність моделювання процесів з довгостроковою пам'яттю [21].

Моделі на базі класифікаційно - регресивних дерев

Переваги моделі на базі класифікаційно - регресивних дерев:

					<i>ДП.ДС71.10.000ПЗ</i>	Арк.
						13
Змн.	Арк.	№ докум.	Підпис	Дата		

- масштабованість, через це відбувається швидка обробка величезних обсягів даних [16];
- можливість використовувати категоріальні зовнішні чинники.

Недоліки [20]:

- неоднозначність алгоритму побудови структури дерева;
- складність питання зупинки, коли варто припинити подальші розгалуження;
- відсутність однаковості їх аналізу та проектування.

1.2 Оцінка прогнозування навантаження

Адекватність моделі - це ступінь її відповідності до реального об'єкту дослідження .

Модель вважають адекватною, якщо систематична компонента (тренд) визначена вірно. При правильному виборі виду тренду, різниці між фактичними рівнями ряду і їх розрахунковими будуть [11]:

1. носить мимовільний характер (зміна залишкової компоненти не пов'язано зі зміною тимчасового періоду);
2. матиме математичне очікування, що дорівнює нулю;
3. нормальний розподіл;
4. незалежні один від одного .

Достовірність моделі характеризується рівнем близькості розрахункових рівнів ряду до фактичних значень. Слід зазначити , що процес оцінки якості прогнозу починається з перевірки якості вихідних даних, коректності теоретичних методик прогнозу, що представляють основу моделі прогнозу.

					<i>ДП.ДС71.10.000ПЗ</i>	Арк.
						14
Змн.	Арк.	№ докум.	Підпис	Дата		

Якість вихідної інформації, залежить від точність вимірювань, діапазону досліджуваних даних, обраного методу дослідження, точність одержаних результатів.

Критерії оцінки якості прогнозу, застосовуються для отримання характеристик, таких як: надійність, точність, достовірність, помилка прогнозу.

Надійність розрахунків - це міра поведінки об'єкта прогнозування в часі. Вона обумовлюється ймовірністю реалізації прогнозу для довірчого інтервалу. Точність прогнозу визначає інтервальний розподіл прогнозних траєкторій при фіксованому рівні достовірності [11].

Побки прогнозування – це відмінності прогнозних значень від реальних значень прогнозованого об'єкта. Але визначити такі параметри як надійність, точність, достовірність, визначити помилки прогнозу не є можливим, оскільки прогнозовані результати немає з чим порівнювати.

З цього випливає, що з'являється проблема оцінки якості прогнозу ще до реалізації.

Якість вихідної інформації, в свою чергу, обумовлюється:

- точністю вимірювань;
- якістю вибірки;
- відсутністю похибок узгодження

Похибки узгодження з'являються тоді, коли вихідна інформація для проведення прогнозних розрахунків оцінюється різними фахівцями, які використовують різні методологічні підходи [12].

Похибки пов'язані з обраною моделлю прогнозу, які з'являються в результаті спрощення або недосконалості теоретичних побудов.

Прогноз формалізованим методом, найчастіше проявляється кількістю показників, яким дана точкова(величини позбавлені змісту, через нульову ймовірність) або інтервальна оцінка.

Інтервальна оцінка - це числовий інтервал в якому, ймовірно, знаходиться прогнозний параметр [12].

					<i>ДП.ДС71.10.000ПЗ</i>	Арк.
						15
Змн.	Арк.	№ докум.	Підпис	Дата		

Достовірність прогнозу залежить від величини похибки, яка розраховується залежністю між прогнозованими і фактичними значеннями досліджуваної величини. Вихідні дані можливо оцінити ще на етапі збору інформації, як ступінь наближених результатів вимірювань до істинного значення вимірюваної величини.

1.3 Похибки прогнозування

Важлива оцінка якості прогнозів – це визначення причин похибок в прогнозуванні.

Найбільш поширені причини похибок прогнозування [13]:

- ігнорується імовірнісний характер прогнозної інформації, вирішенням цієї проблеми є можливим при створенні умовних прогнозів;
- вимоги поставленої точності та достовірності прогнозу, і складності поставленого завдання та глибиною прогнозування;
- недостатня розробленість теорії, на якій базується прогноз [13]. Якісне прогнозування неможливе без теоретичного огляду прогнозованих процесів, оцінка логіки їх протікання, виявлення взаємозв'язків і взаємозалежності;
- складність прогнозування нерегулярних подій, що протікають без певних закономірностей;
- помилки вхідних даних, через не точність вимірювань. Також оцінку якості прогнозу необхідно починати з перевірки вхідних даних, та проаналізувати теоретичний і методичний підхід, що і є основою розробки прогнозування .

Точність прогнозування оцінюється шляхом порівняння фактичних і передбачених значень. Точності прогнозу відбувається за допомогою таких величини, як:

- середньо квадратична похибка прогнозу [14],
- середня абсолютна похибка прогнозу у відсотках,

					ДП.ДС71.10.000ПЗ	Арк.
						16
Змн.	Арк.	№ докум.	Підпис	Дата		

- середня абсолютна похибка прогнозу.

Абсолютні похибки прогнозування, базуються на розрахунку абсолютних похибок і включає в себе оцінки, засновані на обчисленні величини ε_t

$$\varepsilon_t = (y_t - f_t^m), \quad (1.7)$$

де y_t - виміряне значення в момент часу t , f_t^m – прогнозне значення в момент часу t .

Середня абсолютна похибка прогнозу (MAE):

$$MAE = \frac{\sum_{i=1}^N \Delta x_i}{N}. \quad (1.8)$$

MAE - це середня абсолютна величина похибок прогнозу. Цю міру краще використовувати, коли ціна похибки прогнозу пропорційна абсолютній величині цієї похибки. Показник чутливий до одиниць вимірів, тому його не використовують для порівняння якості прогнозу, отриманих з даних які були виміряні в різних одиницях.

Середньоквадратична похибка (Mean squared error, MSE):

$$MSE = \frac{\sum_{i=1}^N \Delta x_i^2}{N}. \quad (1.9)$$

MSE - це функція ризику, що відповідає очікуваному значенню квадратних втрат помилок. Той факт, що MSE майже завжди є суто позитивним (а не нульовим), зумовлений випадковістю або тим, що оцінюване враховує інформацію, яка може дати більш точну оцінку. Він завжди невід'ємний, і значення ближчі до нуля є кращими [14].

Корінь середньоквадратичної похибки (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^N \Delta x_i^2}{N}}. \quad (1.10)$$

					ДП.ДС71.10.000ПЗ	Арк.
						17
Змн.	Арк.	№ докум.	Підпис	Дата		

RMSE - це корінь квадратний із середнього арифметичного квадрата похибок прогнозу. Показник доречний у випадках, коли ціна похибок росте ,як квадрат її значення. Найбільш розповсюджене використання методу –це квадратна функція втрат. Вплив кожної помилки на RMSE пропорційний розміру квадратичної помилки; таким чином більші помилки мають непропорційно великий вплив на RMSE.

Похибки прогнозування у відсотках

Середня абсолютна похибка у відсотках (MAPE):

$$MAPE = \frac{\sum_{i=1}^N \left| \frac{\Delta x_i}{x_{\phi i}} \right|}{N} \cdot 100\%. \quad (1.11)$$

MAPE - це середнє арифметичне абсолютне значення відсотків ,які становлять похибки прогнозу відносно деякого значення. Перевага такого методу, його безрозмірність. Основний мінус – можливість нульового значення для бази, відносно якого розраховується відсоток, а значить, невизначеність значення показника, і крім того великі значення помилок .

Симетричні похибки прогнозування

Критерії, що входять в дану групу, розраховуються на підставі значень s_t , що обчислюється за формулою:

$$S_t = \frac{e_t}{y_t + f_t}. \quad (1.12)$$

Симетрична середня абсолютна процентна похибка:

$$sMAPE = \frac{1}{n} \sum_{i=1}^n 200 \cdot |s_i|. \quad (1.13)$$

Масштабовані помилки прогнозування

В якості базису для розрахунку масштабованих помилок використовується величина q_t .

					ДП.ДС71.10.000ПЗ	Арк.
						18
Змн.	Арк.	№ докум.	Підпис	Дата		

$$q_t = \frac{|e_t|}{\frac{1}{n-1} \sum_{i=2}^n |y_i - y_{i-1}|}. \quad (1.14)$$

До цієї групи входить середня абсолютна масштабована помилка (Mean Absolute Scaled Error, MASE) [2]:

$$\text{MASE} = i_{\substack{\text{maen} \\ = 1, n}} \cdot |q_i|. \quad (1.15)$$

Медіана абсолютної масштабованої помилки знаходиться наступним чином:

$$\text{MdSE} = i_{\substack{\text{median} \\ = 1, n}} \cdot |q_i|. \quad (1.16)$$

Середньоквадратична масштабована помилка (Root Mean Square Scaled Error, RMSSE), що обчислюється за формулою [3]:

$$\text{RMSSE} = \sqrt{i_{\substack{\text{median} \\ = 1, n}} \cdot (|q_i|)^2}. \quad (1.17)$$

Відносні оцінки

Відносна середня абсолютна помилка:

$$\text{RMAE} = \frac{\text{MAE}}{\text{MAE}^*}, \quad (1.18)$$

де MAE і MAE * середні абсолютні помилки для аналізованої моделі прогнозування і базової моделі відповідно.

Відносне середньоквадратичне відхилення (Relative Root Mean Square Error, RelRMSE) обчислюється аналогічно

$$\text{RMAE} = \frac{\text{RMSE}}{\text{RMSE}^*}. \quad (1.19)$$

Оцінка логарифм відносної середньоквадратичного відхилення (Log mean squared error ratio, LMR):

$$\text{LMR} = \log\left(\frac{\text{RMSE}}{\text{RMSE}^*}\right). \quad (1.20)$$

					ДП.ДС71.10.000ПЗ	Арк.
						19
Змн.	Арк.	№ докум.	Підпис	Дата		

Наступна група оцінок розраховує число випадків, коли похибка досліджуваної моделі прогнозування перевищує значення похибки базової моделі.

До таких оцінок відносяться РВ (MAE) - Percentage Better (MAE),

$$PB(MAE) = 100\% \cdot \text{mean}(I\{MAE < MAE^*\}). \quad (1.21)$$

Отже, похибки прогнозування, важлива оцінка якості прогнозованої моделі. Для оцінки формалізованих методів, а саме часових рядів, використовують середню абсолютну похибку.

					ДП.ДС71.10.000ПЗ	Арк.
						20
Змн.	Арк.	№ докум.	Підпис	Дата		

2. ПОБУДОВА ПРОГНОЗНОЇ МОДЕЛІ

Розглянувши методи прогнозування навантаження у MicroGrid найбільш придатним є метод регресійного аналізу, оскільки поведінка кривої навантаження описується функцією часу, тому обираємо моделі часових рядів.

Основна задача прогнозної моделі полягає в прогнозуванні добового значення навантаження, спираючись на данні погодних умов і попереднього спостереження (залежності навантаження від зовнішнього фактору). Вхідні параметрами прогнозної моделі: добові значення погодинного навантаження, параметри зовнішніх умов.

2.1 Добові криві зміни потужності навантаження

Для знаходження статистичної залежності між зовнішнім середовищем та навантаженням приватного будинку, необхідно побудувати погодинну реєстрацію навантажень, для цього розглянемо режими електропостачання, і розрахуємо значення навантаження за добу.

Помірний режим електропостачання передбачає помітні зміни в способі життя при збереженні високого рівня комфорту. Втім, список споживачів мало відрізняється від режиму комфортного енергопостачання, за винятком таких необов'язкових елементів, як електро-підігрів підлоги. Використання електропідігрівача гарячої води теж може бути обмежена. Крім цього, зміни стосуються і часу виконання не надто регулярних, але енергоємних робіт. До цих робіт, наприклад, відноситься прання, в машині-автоматі з підігрівом води, прасування, активна робота з потужним електроінструментом і садової електротехнікою. Складемо табл. 2.1 орієнтовного навантаження будинку.

					ДП.ДС71.10.000ПЗ	Арк.
						21
Змн.	Арк.	№ докум.	Підпис	Дата		

Виділяється дві категорії регулярних споживачів. До першої відносим ті, які важко або некомфортно замінити ручною працею, або виконати цю роботу за допомогою пристроїв, що не вимагають електрики. До другої віднесені ті, що в принципі, в даний момент, працюють на електриці, але досить легко замінюються на неелектричні аналоги. В окрему групу виділено епізодичні споживачі електроенергії, роботу більшості з яких цілком можна замінити ручною працею.

Таблиця 2.1

Орієнтовне електричне навантаження приватного будинку

Споживач	Потужність	Сезон	Проміжок роботи за добу(кВт · год)		Споживання за добу(кВт · год)	
			середнє	максимум	середнє	максимум
Регулярні споживачі						
Освітлення (одночасно 5 енергозберігаючи х ламп по 20 Вт)	0,1	зима	8	10	0,8	1
		літо	1	3	0,1	0,3
Холодильник	0,5	зима	1,5	2	0,75	1
		літо	2	3	1	1,5
Насос вібраційний	0,25	зима	0,5	0,67	0,13	0,17
		літо	1	3	0,25	0,75
Насос відцентровий	0,8	постійно	0,25	0,5	0,2	0,4
Пральна машина	0,5	постійно	0,5	3	0,25	1,5
Праска	1,5	постійно	0,17	1	0,25	1,5
Телевізор	0,15	постійно	2	4	0,3	0,6
Ноутбук	0,1	постійно	2	4	0,2	0,4
Всього	3,9кВт	зима			2,88	6,57
		літо			2,55	6,95

Не регулярні споживачі						
Кухонні прибори (кухонний комбайн, мясорубка, міксер)	2	постійно	0,25	2	0,5	4
Косметичні електроприбори (бритва, фен)	2	постійно	0,08	0,33	0,17	0,67
Пилосос	1,8	постійно	0,33	1	0,6	1,8
Електроінструмен т (болгарка, дріль, лобзик, електропилка)	2	постійно	0,5	4	1	8
Газонокосарка	1,5	літо	1	3	1,5	4,5
Снігоприбиральни к	1,5	зима	1	3	1,5	4,5
Всього	10,8 кВт	зима			3,77	18,97
		літо			3,77	18,97
Всього по будинку	14,7 кВт	зима			6,64	25,53
		літо			6,32	25,92

Для споживачів окремо виділимо роботу зимою та літом, оскільки різний період року впливає на їх використання. При цьому, якщо для оцінювання зимового режиму слід орієнтуватися на дні зимового сонцестояння (21 – 23 грудня) або дні з найбільшими морозами (зазвичай друга половина січня). А для оцінювання літнього режиму необхідно брати теплий період з найбільш довгими та теплими днями, це кінець липня або початок серпня.

Підсумок, для помірного режиму. Якщо виключити регулярних споживачів другої черги (чайник і водонагрівачі), то слід орієнтуватися на щомісячне споживання близько 210 кВт·год при миттєвої споживаної потужності в тривалому режимі до 3 кВт і пікової потужності до 7 кВт, а очікуване середньодобове навантаження становить 4 - 6 кВт·год.

Максимум навантаження за тиждень припадає на суботу, це пояснюється тим, що в цей день виконуються різні роботи по дому.

Мінімум – припадає на один з робочих днів (це залежить від стилю життя сім'ї, що проживає в приватному будинку). Графік навантаження електроспоживання за тиждень (рис. 2.1).

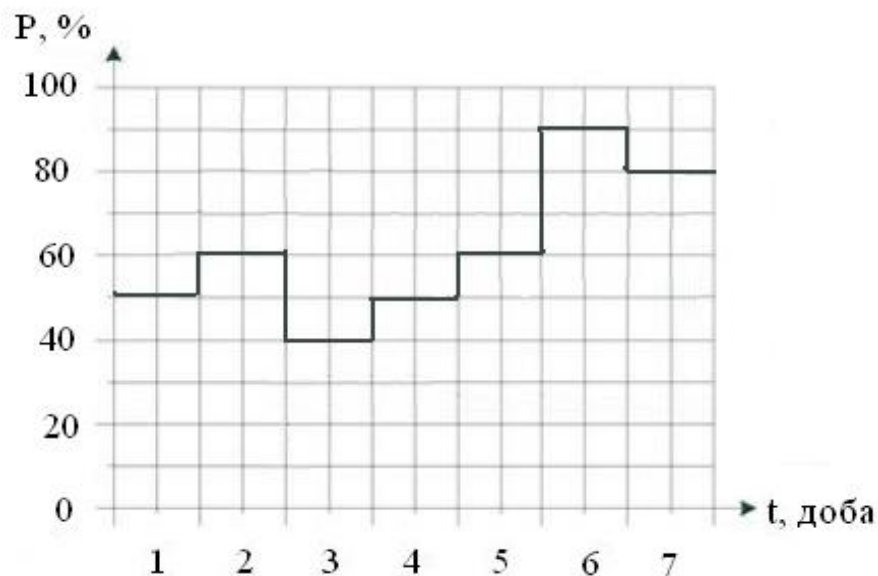


Рис. 2.1. Графік навантаження приватного будинку за тиждень

Період максимального навантаження в приватному будинку з 17.00 до 21.00 години та 6:00 до 9:00. Період мінімального навантаження – з 12.00 до 13.00 години та частково з 22:00 до 04:00 (рис. 2.2).

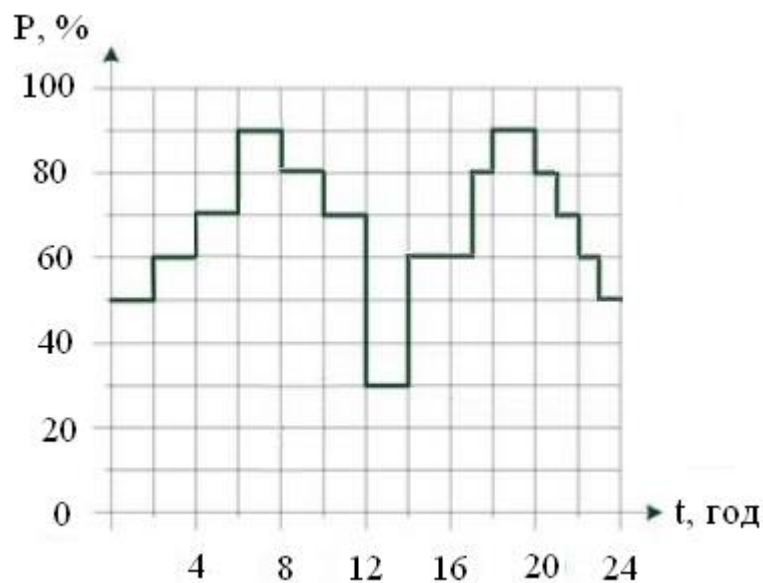


Рис. 2.2. Добовий графік навантаження приватного будинку робочий день

Змн.	Арк.	№ докум.	Підпис	Дата

ДП.ДС71.10.000ПЗ

Арк.

24

Заповнимо табл. 2.2 згідно процентного навантаження, та обчисленого раніше, середньодобове споживання за добу 3-4 кВт.

Таблиця 2.2

Розподіл використаного навантаження по годинах

P, %	P, кВт	t, год	P, %	P, кВт	t, год
15	3,5	0	64	2,1	13
14	3,5	1	55	2,1	14
12	4,2	2	50	4,2	15
10	4,2	3	80	4,2	16
10	4,9	4	82	4,2	17
16	4,9	5	81	5,6	18
36	6,3	6	90	5,6	19
40	6,3	7	85	6,3	20
60	5,6	8	77	5,6	21
65	5,6	9	45	4,9	22
66	4,9	10	28	4,2	23
70	4,9	11	64	3,5	24
65	3,5	12			

Таким чином маючи графік навантаження одного робочого дня і значення навантаження за добу ,можемо створити графік 5-ти робочих днів змінюючи відсоток погодинного навантаження на 5% .

Отримані добові графіки навантаження за робочі дні можемо спостерігати на (рис.2.3).

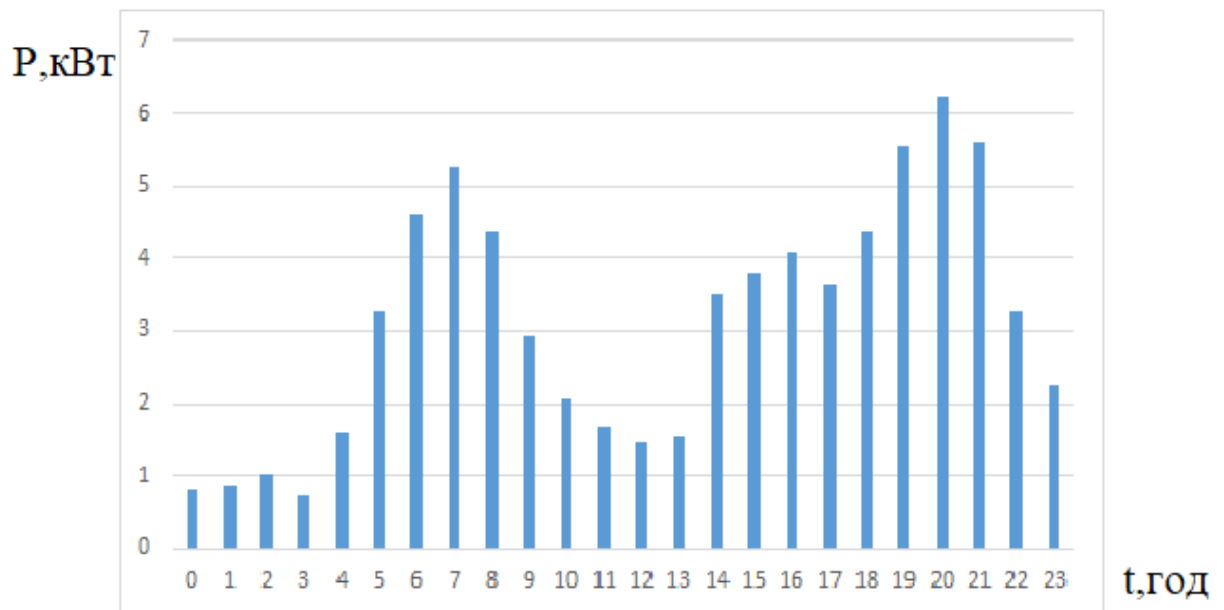


Рис. 2.3. Добовий графік навантаження (Понеділок)

Навантаження робочих днів відрізняються від навантаження вихідного дня. Тому беремо графік добового навантаження (вихідного дня) приватного будинку і відсоткове розподілення навантаження за кожну годину (рис.2.4). Будуємо імітацію погодинної реєстрації даних навантаження за вихідний день (рис.2.5).

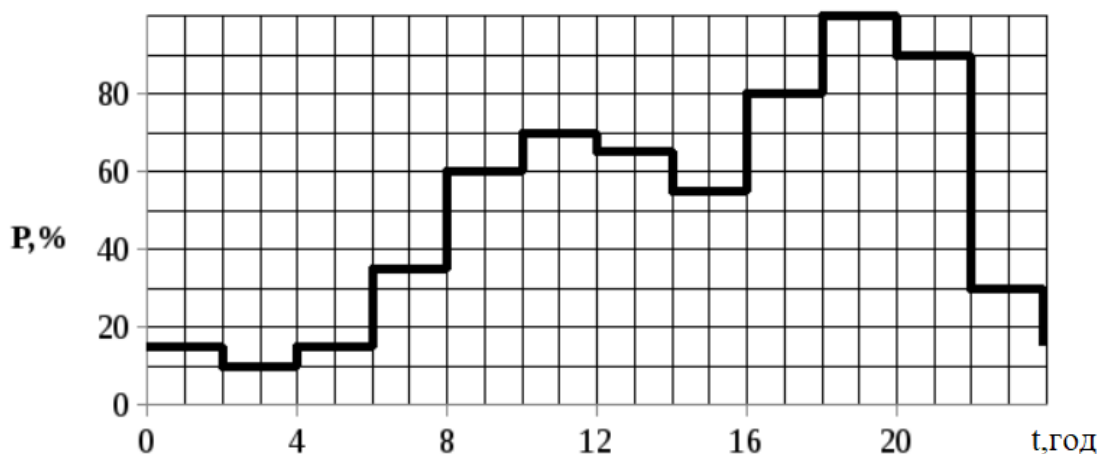


Рис. 2.4. Добовий графік навантаження приватного будинку (вихідний день)

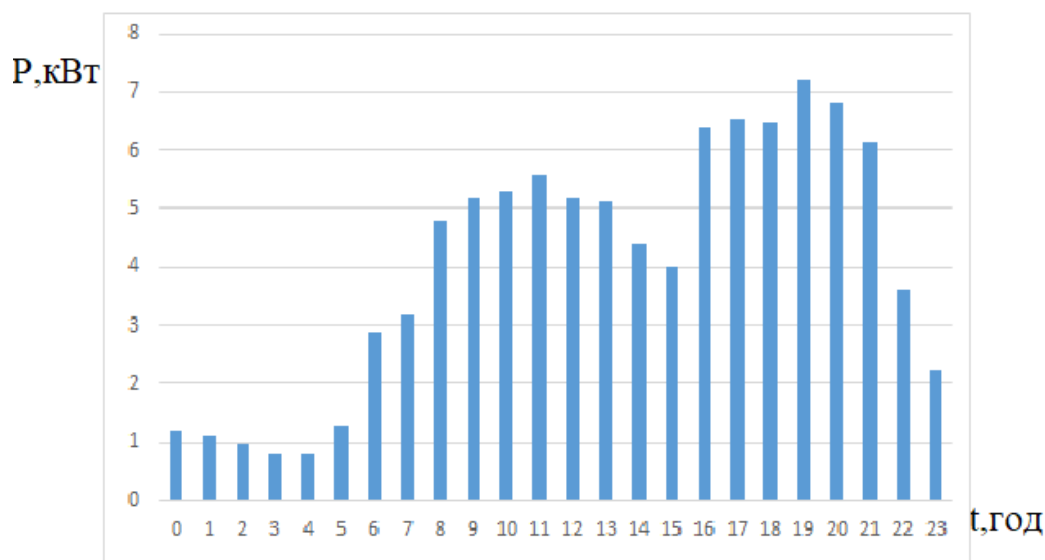


Рис. 2.5. Добовий графік навантаження (Субота)

Отже, використовуючи графіки добових навантажень (робочого і вихідного дня) приватного будинку і відсоткове розподілення навантаження за кожну годину, побудували імітацію погодинної реєстрації даних навантаження за кожен день тижня.

Модель щогодинної реєстрації зовнішніх погодних умов

Для знаходження залежність між навантаженням і навколишнім середовищем, обрано такі фактори як: температура, тиск, вологість. За допомогою інтернет ресурсу meteo.ua, отримали данні щогодинних значень: температури, тиску, вологості, приведемо їх в табл. 2.3.

Таблиця 2.3

Щогодинні значення за 5.04.20

час	температура	тиск	вологість
0	7	755	21
1	2	755	37
2	0	755	47
3	5	756	27
4	0	756	47
5	0	756	47

Продовження табл. 2.3

6	3	757	34
7	-1	757	51
8	2	758	52
9	6	759	31
10	7	759	37
11	9	759	32
12	9	759	23
13	11	759	20
14	11	759	18
15	11	759	18
16	12	758	19
17	11	758	20
18	11	758	18
19	10	758	23
20	6	759	28
21	9	760	24
22	6	759	28
23	3	759	38

Використовуючи дані з таблиці 2.2 і таблиці 2.3, перевіримо чи є статистичний зв'язок між параметрами навантаження і параметрами навколишнього середовища, розрахувавши коефіцієнти регресії отримали:

- між навантаженням і температурою $r=0,72$, це свідчить про високий позитивний зв'язок між даними дослідження ,
- між навантаженням і тиском $r=0,76$, це свідчить про високий позитивний зв'язок між даними дослідження ,
- між навантаженням і тиском $r=-0,51$, це свідчить про середню обернену залежність між даними.

Отже ,спираючись на результати розрахованих коефіцієнтів кореляції, робимо висновок ,що обрані дані дослідження є статистично залежними.

					ДП.ДС71.10.000ПЗ	Арк.
						28
Змн.	Арк.	№ докум.	Підпис	Дата		

Ця залежність дозволяє створювати прогнозу модель, спираючись на результат попередніх спостережень .

2.2 Оцінка подібності днів за даними навколишнього середовища

Достовірність прогнозу залежить від величини похибки, яка розраховується залежністю між прогнозованими і фактичними значеннями досліджуваної величини. Вхідні дані можливо оцінити ще на етапі збору інформації, як ступінь наближених результатів вимірювань до істинного значення вимірюваної величини.

Тобто, щоб збільшити точність прогнозу моделі необхідно зібрати достатню кількість даних, щоб сформував базу. Зібравши достатній об'єм даних, необхідно виконати сортування за деякими критеріями. Тому створюємо базу даних в якій будуть зберігатися дані навколишнього середовища які групуються за розрахованими критеріями (мат очікування та дисперсії).

Маємо поточні дані навколишнього середовища за вихідний день ,а саме неділю (табл. 2.3), значення яких задані на інтервалі від 00:00 до 23:00. Для того, щоб спрогнозувати навантаження на цю конкретну добу, потрібно знайти в базі найбільш схожих два дні. Для цього розраховуємо мат очікування і дисперсії отриманих даних навколишнього середовища. Мат очікування та дисперсія розраховується за формулами :

$$E(\xi) = \int_X f(x) dF_{\xi}(x). \quad (2.1)$$

$$D(x) = M(x^2) - (M(x))^2. \quad (2.2)$$

Розрахувавши мат очікування та дисперсію заданого дня, отримані дані занесемо в табл. 2.4

					ДП.ДС71.10.000ПЗ	Арк.
						29
Змн.	Арк.	№ докум.	Підпис	Дата		

Таблиця 2.4

Розраховані параметри дисперсії та маточікування

	Температура	Тиск	Вологість
Мат очікування	6,25	757,7917	30,83333
Дисперсія	17,1875	2,248264	122,6389

На основі цих даних необхідно зробити пошук в базі даних навколишнього середовища, і знайти два схожих дня, для подальшого прогнозування.

В ході дослідження була встановлена максимальна допустима відсоткова різниця між значеннями заданого і схожого дня, яка становить 14%. Не перевищуючи 14%, прогнозне значення навантаження є адекватним при порівнянні прогнозованого і реального навантаження.

Здійснивши пошук за критеріями:

- день тижня (робочий, вихідний);
- мат очікування;
- дисперсія;

Було знайдено два схожих дні, заповнимо табл. 2.5 та табл. 2.6.

Таблиця 2.5.

Параметри схожого дня за даними навколишнього середовища

	Температура	Тиск	Вологість
Мат очікування	6,791667	746,7083333	35
Дисперсія	16,0816	2,039930556	138,5

Таблиця 2.6

Параметри другого схожого дня за даними навколишнього середовища

	Температура	Тиск	Вологість
Мат очікування	6,208333	744,4583	34,83333
Дисперсія	18,0816	2,164931	122,8056

					<i>ДП.ДС71.10.000ПЗ</i>	Арк.
						30
Змн.	Арк.	№ докум.	Підпис	Дата		

Отримавши статистично схожі дні, наступним кроком буде побудова регресивного рівняння.

2.3 Побудова прогнозової моделі за регресивними рівняннями

Для отриманих подібних днів розраховуємо регресивні рівняння. Рівняння регресії може бути представлено у вигляді:

$$Y = f(\beta, X) + \epsilon, \quad (2.3)$$

де $X = (X_1, X_2, X_3 \dots X_m)$ – залежні змінні; β - вектор параметрів; ϵ - випадкове відхилення; Y - залежна змінна.

На прогнозну модель впливають три фактори, тому використовуємо рівняння множинної регресії. Рівняння множинної регресії має вигляд:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_m X_m + \epsilon, \quad (2.4)$$

Беремо перший знайдений схожий день і розраховуємо регресивні коефіцієнти та перевіряємо їх на значимість, щоб отримати регресивне рівняння.

1.Рівняння регресії

Розрахуємо коефіцієнти регресії. За допомогою методу найменших квадратів, вектор S виходить з виразу:

$$S = (X^T X)^{-1} X^T Y. \quad (2.5)$$

Матриця Y в табл. 2.7.

Таблиця 2.7

Значення навантаження

1.095	0.74	4.38	4.745	5.84	6.205
1.022	1.168	4.745	4.672	5.986	5.621
0.876	2.628	4.818	4.015	5.913	3.285
0.73	2.92	5.11	3.65	6.57	2.044

Матриці зі змінними X_j (даними навколишнього середовища) додаємо одиничний стовпець і записуємо в табл. 2.8

					ДП.ДС71.10.000ПЗ	Арк.
						31
Змн.	Арк.	№ докум.	Підпис	Дата		

Таблиця 2.8

Матриця даних навколишнього середовища

1	4	754	32
1	1	752	40
1	0	752	41
1	0	752	35
1	1	752	45
1	2	752	50
1	2	752	42
1	3	752	56
1	5	754	48
1	6	754	45
1	7	754	37
1	8	754	35
1	9	756	24
1	10	755	23
1	10	756	23
1	12	756	20
1	13	756	19
1	11	755	19
1	10	755	20
1	10	755	20
1	9	755	21
1	8	754	21
1	8	754	21
1	6	752	22

Множимо матриці, $(X^T X)$. У матриці, $(X^T X)$ число 24, що знаходиться на перетині першого рядка і стовпчика, розрахуємо сума добутків елементів першого рядка матриці X^T і першого стовпця матриці X .

Перемножуємо матриці, $(X^T Y)$, знаходимо обернену матрицю $(X^T X)^{-1}$.

Вектор оцінок коефіцієнтів регресії:

$$Y(X) = \begin{vmatrix} 66354.025 & 31.328 & -88.3 & 0.334 \\ 31.328 & 0.022 & -0.0418 & 0.00207 \\ -88.3 & -0.0418 & 0.118 & -0.000495 \\ 0.334 & 0.00207 & -0.000495 & 0.000813 \end{vmatrix} \cdot \begin{vmatrix} 88.778 \\ 726.798 \\ 66980.202 \\ 2507.197 \end{vmatrix} = \begin{vmatrix} 58.128 \\ 0.554 \\ -0.0789 \\ 0.0472 \end{vmatrix} \quad (2.6)$$

Розрахувавши коефіцієнти регресії отримали рівняння:

$$Y = 58.1281 + 0.5544X_1 - 0.07893X_2 + 0.04717X_3 \quad (2.7)$$

2. парні коефіцієнти кореляції R

Розрахуємо парні коефіцієнти кореляції, за формулою.

$$r_{xy} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{s(x) \cdot s(y)}. \quad (2.8)$$

Значення парного коефіцієнта кореляції свідчить про сильний зв'язок між x_1 і y , де x_1 - температура, y - навантаження.

$$r_{yx_1} = 0.841. \quad (2.9)$$

Коефіцієнт кореляції між x_2 і y , де x_2 - тиск, y - навантаження, свідчить про сильний зв'язок.

$$r_{yx_2} = 0.764. \quad (2.10)$$

Значення парного коефіцієнта кореляції свідчить про помірний зв'язок між x_3 та y , де x_3 - вологість, y - навантаження.

$$r_{yx_3} = -0.555. \quad (2.11)$$

Значення парного коефіцієнта кореляції свідчить про досить сильний зв'язок між x_2 і x_1 (між тиском і температурою).

$$r_{x_1 x_2} = 0.915. \quad (2.12)$$

Коефіцієнт кореляції свідчить про сильний зв'язок між x_3 та x_1 (між вологістю і температурою) .

$$r_{x_1 x_3} = -0.787. \quad (2.13)$$

Парний коефіцієнт кореляції свідчить про сильний лінійний зв'язок між x_3 та x_2 (між вологістю і тиском).

$$r_{x_2 x_3} = -0.708. \quad (2.14)$$

Розрахуємо і заповнимо в табл. 2.9 значення дисперсії та середньо квадратичного відхилення.

Таблиця 2.9

Значення дисперсії та середньо квадратичного відхилення

Признаки x и y	$D(x)$	$D(y)$	$S(x)$	$S(y)$
Для y и x_1	155	6.458	88.778	3.699
Для y и x_2	18093	753.875	88.778	3.699
Для y и x_3	759	31.625	88.778	3.699
Для x_1 и x_2	18093	753.875	155	6.458
Для x_1 и x_3	759	31.625	155	6.458
Для x_2 и x_3	759	31.625	18093	753.875

Матриця парних коефіцієнтів кореляції R заповнимо в табл. 2.10:

Таблиця 2.10

Парні коефіцієнти кореляції

-	y	x_1	x_2	x_3
y	1	0.8411	0.7637	-0.5552
x_1	0.8411	1	0.9154	-0.7869
x_2	0.7637	0.9154	1	-0.7077
x_3	-0.5552	-0.7869	-0.7077	1

Розрахуємо спостережувані значення t -статистики для r_{yx_1} по формулі:

$$t_{\text{спост}} = r_{yx_1} \cdot \frac{\sqrt{n-m-1}}{\sqrt{1-r_{yx_1}^2}}, \quad (2.15)$$

де $m = 1$ - кількість факторів в рівнянні регресії.

$$t_{\text{спост}} = 0.84 \cdot \frac{\sqrt{24-1-1}}{\sqrt{1-0.84^2}} = 7.3. \quad (2.16)$$

Використовуємо таблиці Стюдента за допомогою якої знаходимо $t_{\text{крит}} = 2.405$. Оскільки $t_{\text{спост}} > t_{\text{крит}}$, то гіпотеза про рівність нулю кореляції, відпадає. З цього випливає, що кореляційний коефіцієнт значущий.

Розрахуємо спостережувані значення t -статистики для навантаження і тиску r_{yx_2} по формулі:

$$t_{\text{спост}} = 0.76 \cdot \frac{\sqrt{24-1-1}}{\sqrt{1-0.76^2}} = 5.5. \quad (2.17)$$

Оскільки $t_{\text{спост}} > t_{\text{крит}}$, то гіпотеза про рівність нулю кореляції, відпадає. З цього випливає, що кореляційний коефіцієнт значущий.

Розрахуємо t -статистики (спостережені) для навантаження і вологості r_{yx_3} :

$$t_{\text{спост}} = 0.56 \cdot \frac{\sqrt{24-1-1}}{\sqrt{1-0.56^2}} = 3.13. \quad (2.18)$$

Оскільки $t_{\text{спост}} > t_{\text{крит}}$, то гіпотеза про рівність нулю кореляції, відпадає. Отже, коефіцієнт кореляції статистично – значущий.

3. Аналіз коефіцієнтів регресії

Виконаємо аналіз розрахованого регресивного рівняння. Перевіривши на значимість коефіцієнтів регресії, дослідження абсолютних та відносних похибок апроксимації табл. 2.11.

Середня похибка апроксимації:

$$A = \frac{\sum \left| \frac{\varepsilon}{Y} \right|}{n} \cdot 100\% = \frac{7.643}{24} \cdot 100\% = 31.85\%. \quad (2.19)$$

					ДП.ДС71.10.000ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		35

Оцінка середньоквадратичного відхилення (стандартна помилка для оцінки навантаження Y), розраховується

$$S = \sqrt{S^2} = 1.088. \quad (2.20)$$

Таблиця 2.11

Дослідження абсолютних та відносних похибок апроксимації

Y	$Y(x)$	$\varepsilon = Y - Y(x)$	ε^2	$(Y - Y_{cp})^2$	$ E: Y $
1.095	2.344	-1.249	1.56	6.781	1.141
1.022	1.216	-0.194	0.0376	7.167	0.19
0.876	0.709	0.167	0.028	7.97	0.191
0.73	0.426	0.304	0.0927	8.815	0.417
0.74	1.452	-0.712	0.507	8.756	0.962
1.168	2.242	-1.074	1.153	6.406	0.92
2.628	1.865	0.763	0.583	1.147	0.29
2.92	3.079	-0.159	0.0254	0.607	0.0546
4.38	3.653	0.727	0.528	0.464	0.166
4.745	4.066	0.679	0.461	1.094	0.143
4.818	4.243	0.575	0.331	1.252	0.119
5.11	4.703	0.407	0.166	1.991	0.0796
4.745	4.581	0.164	0.0269	1.094	0.0346
4.672	5.167	-0.495	0.245	0.947	0.106
4.015	5.088	-1.073	1.152	0.0998	0.267
3.65	6.055	-2.405	5.786	0.00241	0.659
5.84	6.563	-0.723	0.522	4.584	0.124
5.986	5.533	0.453	0.205	5.23	0.0757
5.913	5.026	0.887	0.788	4.901	0.15
6.57	5.026	1.544	2.385	8.242	0.235
6.205	4.518	1.687	2.845	6.28	0.272
5.621	4.043	1.578	2.491	3.694	0.281
3.285	4.043	-0.758	0.574	0.171	0.231
2.044	3.139	-1.095	1.199	2.739	0.536

4. Множинний коефіцієнт кореляції

За допомогою множинного коефіцієнта кореляції перевіряємо спільний вплив чинників(даних навколишнього середовища) на результат прогнозу.

Чим ближче значення R до одиниці, тим краще рівняння регресії описує фактичні дані і фактори сильніше впливають на результат. Якщо R близький до нуля, то рівняння регресії погано описує фактичні дані і тому фактори чинять слабкий вплив на результат.

$$R = \sqrt{1 - \frac{se^2}{\sum(y - \bar{y})^2}} = \sqrt{1 - \frac{23.691}{90.43}} = 0.8591. \quad (2.21)$$

Коефіцієнт детермінації :

$$R^2 = 0.8591^2 = 0.738. \quad (2.22)$$

Більш об'єктивною оцінкою є скоригований коефіцієнт детермінації:

$$R^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-m-1}. \quad (2.23)$$

$$R^2 = 1 - (1 - 0.738) \cdot \frac{24-1}{24-3-1} = 0.699. \quad (2.24)$$

Отже, цей коефіцієнт добре пояснює поведінку Y(навантаження).

5. Перевірка якості рівняння множинної регресії

Перевірка якості рівняння множинної регресії здійснюється шляхом гіпотези про рівність нулю коефіцієнта детермінації. Для перевірки цієї гіпотези використовують F-критерій Фішера. Тому обчислимо фактичне (що спостерігається) значення F-критерію, за допомогою коефіцієнта детермінації R^2 , який був розрахований за даними спостереження.

Спочатку необхідно, використовуючи таблицю Фішера-Снедекора знайти критичне значення F-критерію ($F_{кр}$). Задаємо рівень значущості α (0,05) і двома числами ступенів свободи $k_1 = m$ і $k_2 = nm - 1$.

					ДП.ДС71.10.000ПЗ	Арк.
						37
Змн.	Арк.	№ докум.	Підпис	Дата		

Розраховуємо Критерій Фішера.

$$R^2 = 1 - \frac{se^2}{\sum (y - \bar{y})^2} = 1 - \frac{23.691}{90.43} = 0.788. \quad (2.25)$$

Необхідно перевірити гіпотезу за допомогою F-статистики розподілу Фішера, якщо $F < F_{кр}$, то немає підстав для відхилення гіпотези.

$$R = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m} = \frac{0.738}{1 - 0.738} \cdot \frac{24 - 3 - 1}{3} = 18.781. \quad (2.26)$$

Розрахуємо табличне значення при ступенях вільності $k_1 = 3$ і $k_2 = nm - 1 = 24 - 3 - 1 = 20$, $F_{кр}(3; 20) = 3.1$. Отримали фактичне значення $F > F_{кр}$, то коефіцієнт детермінації статистично значимий (тобто коефіцієнти b спільно значимі).

В результаті розрахунків було отримано рівняння множинної регресії і перевірено на значимість коефіцієнтів регресії першого дня :

$$Y_1 = 58.1281 + 0.5544x_1 - 0.07893x_2 + 0.04717x_3. \quad (2.27)$$

Спираючись на результати перевірок критерію Фішера та коефіцієнта детермінації, робимо висновок про статистичну значимість рівняння і встановлено, що параметри моделі статистично значимі. Таким самим чином було розраховано і перевірено на значимість та отримано рівняння множинної регресії другого обраного дня :

$$Y_2 = 1.691988 - 0.552461x_1 - 0.00465x_2 + 0.0629346x_3 \quad (2.28)$$

Маючи регресивні рівняння двох схожих днів , розрахуємо першу кінцеву різницю між коефіцієнтами регресії , віднявши від другого рівняння перше. Отриману різницю додаємо до коефіцієнтів регресії другого рівняння і записуємо до табл. 2.12 .

					ДП.ДС71.10.000ПЗ	Арк.
						38
Змн.	Арк.	№ докум.	Підпис	Дата		

Таблиця 2.12

Коефіцієнти регресивних рівнянь

Коеф.Регрес	Y1	Y2	Y2-Y1	Y3=Y2+(Y2+Y1))
b_0	58,1281	1,691988	-56,4454	-54,7534
b_1	0,5544	0,552461	-0,00211	0,550349
b_2	-0,07893	-0,00465	0,074289	0,069637
b_3	0,04717	0,062935	0,015768	0,078703

Розраховані коефіцієнти записуємо у формулу :

$$Y_3 = 54.7534 - 0.550349x_1 + 0.069637x_2 + 0.078703x_3. \quad (2.29)$$

Використовуючи отримане регресивне рівняння, розрахуємо прогнозоване навантаження за 5.04.20. Для цього підставимо у (2.29) дані навколишнього середовища (X1-температура, X2-тиск, X3-вологість) з табл. 2.3. Та отримаємо прогнозоване добове навантаження табл. 2.13.

Таблиця 2.13

Прогнозоване добове навантаження

час	Прогнозоване навантаження	Різниця між реальним та прогнозованим навантаженням	Середня абсолютна похибка	Похибка відносно до середнього значення навантаження
0	3,327680662	-1,88768	1,310889	0,464565
1	1,835178798	-0,71518	0,638552	0,176008
2	1,521507636	-0,56151	0,584904	0,138189
3	2,768835767	-1,96884	2,461045	0,484537
4	1,591144551	-0,79114	0,988931	0,194703
5	1,591144551	-0,31114	0,243082	0,076574
6	2,288693534	0,591306	0,205315	0,145523
7	1,425243223	1,774757	0,554611	0,436774
8	3,224629475	1,575371	0,328202	0,387704
9	3,842906053	1,357094	0,26098	0,333985
10	4,865470917	0,414529	0,078509	0,102017
11	5,57265539	0,027345	0,004883	0,00673
12	4,864331431	0,335669	0,064552	0,082609

Змн.	Арк.	№ докум.	Підпис	Дата

ДП.ДС71.10.000ПЗ

Арк.

39

Продовження табл. 2.13

13	5,728921228	-0,60892	0,11893	0,149858
14	5,571515903	-1,17152	0,266254	0,288314
15	5,571515903	-1,57152	0,392879	0,386755
16	6,130930542	0,269069	0,042042	0,066219
17	5,659284312	0,900716	0,137304	0,221669
18	5,501878988	0,978121	0,150945	0,240719
19	5,345043407	1,854957	0,257633	0,456511
20	3,606798067	3,193202	0,469589	0,785858
21	5,012671008	1,147329	0,186255	0,282362
22	3,606798067	-0,0068	0,001888	0,001673
23	2,742778013	-0,50278	0,224454	0,123735
			41,55%	25,14%

Таким чином прогнозований графік має наступний вигляд (рис. 2.6.)

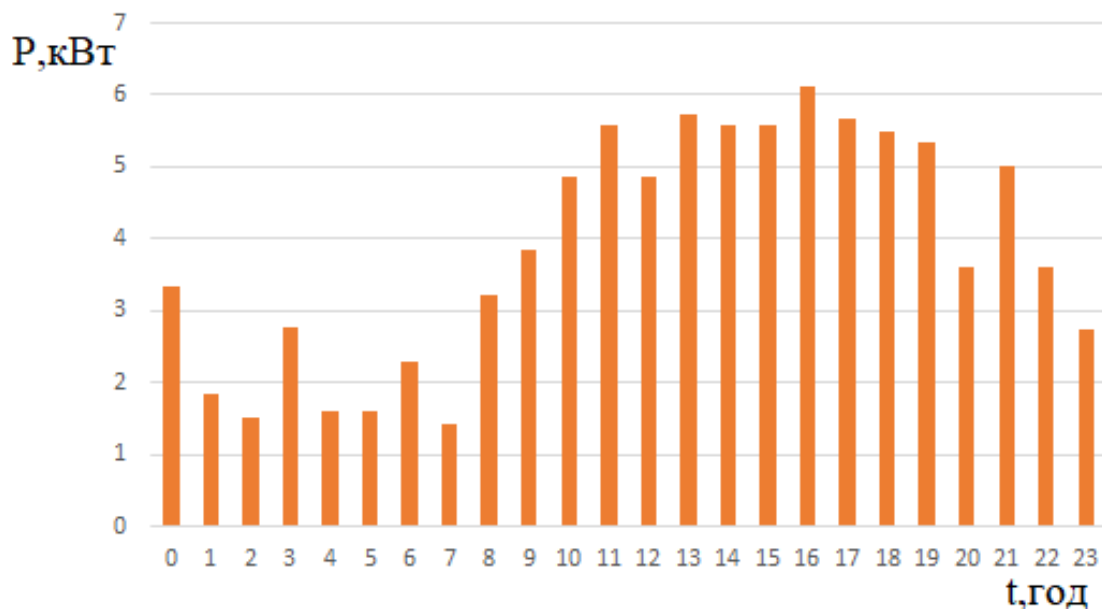


Рис 2.6 Прогнозований графік навантаження

Порівняємо реальний графік навантаження та прогнозований на (рис.2.7.):

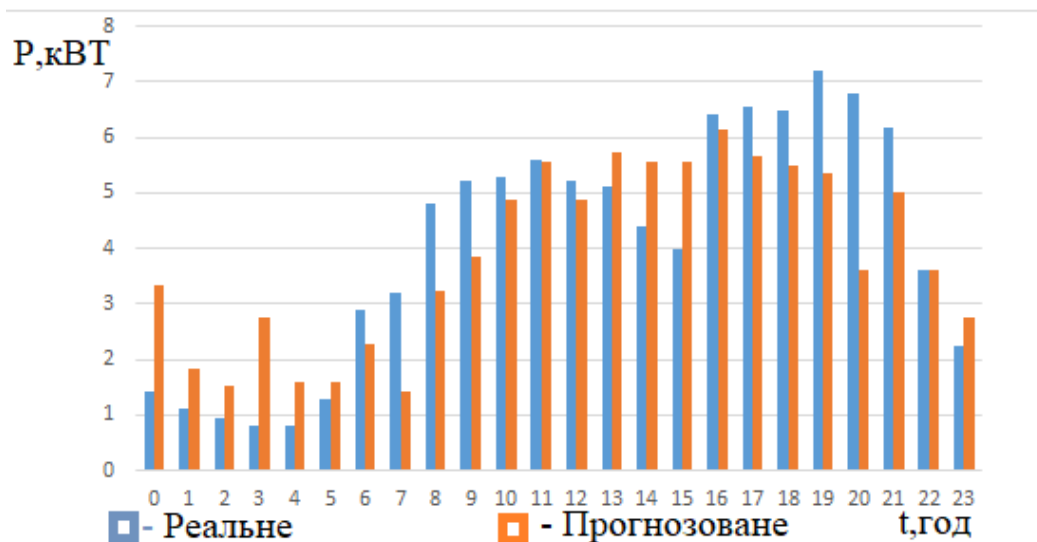


Рис 2.7 Порівняльний графік навантаження

Маючи дані реального і прогнозованого навантаження ,розрахуємо похибку прогнозовної моделі за формулою середньої абсолютної похибки MAPE.

$$MAPE = \frac{\sum_{i=1}^N \left| \frac{\Delta X_1}{X_{\phi i}} \right|}{N} = 41.5\%. \quad (2.30)$$

Похибка прогнозування складає 41%.

3. ПРОГРАМНА РЕАЛІЗАЦІЯ ПРОГНОЗУВАННЯ НАВАНТАЖЕННЯ

3.1 Опис алгоритму

Загальний алгоритм програмної реалізації прогнозування погодинного навантаження в обраний день складається з наступних етапів:

- 1) Завантаження погодинного архіву погоди в Києві.
- 2) Вибір необхідних параметрів погоди, від залежить навантаження.
- 3) Розрахунок необхідних статистик параметрів за кожну добу, яка міститься в архіві погоди.
- 4) Вибір дати, за яку потрібно отримати прогноз навантаження.
- 5) Вибір методу пошуку дат, які найбільше схожі (за погодою) на обрану дату. При цьому враховуються лише статистики, обрані на етапі 3.
- 6) Вибір граничної допустимої розбіжності між датами за обраними статистиками. Якщо розбіжність буде більшою, то дати будуть вважатися несхожими.
- 7) Обчислення двох найближчих дат до дати, обраної на етапі 4, методом, обраним на етапі 5, з урахуванням обмежень, обраних на етапі 6.
- 8) Вибір методу прогнозування навантаження на основі результатів, отриманих на етапі 7.
- 9) Обчислення прогнозу погодинного навантаження на дату, обрану на етапі 4, на основі погоди у дві дати, обчислені на етапі 7, методом, обраним на етапі 8.
- 10) Побудова графіку навантаження, аналіз результатів.

					ДП.ДС71.10.000ПЗ	Арк.
						42
Змн.	Арк.	№ докум.	Підпис	Дата		

Схема алгоритму програми, наведена на рис. 3.1.

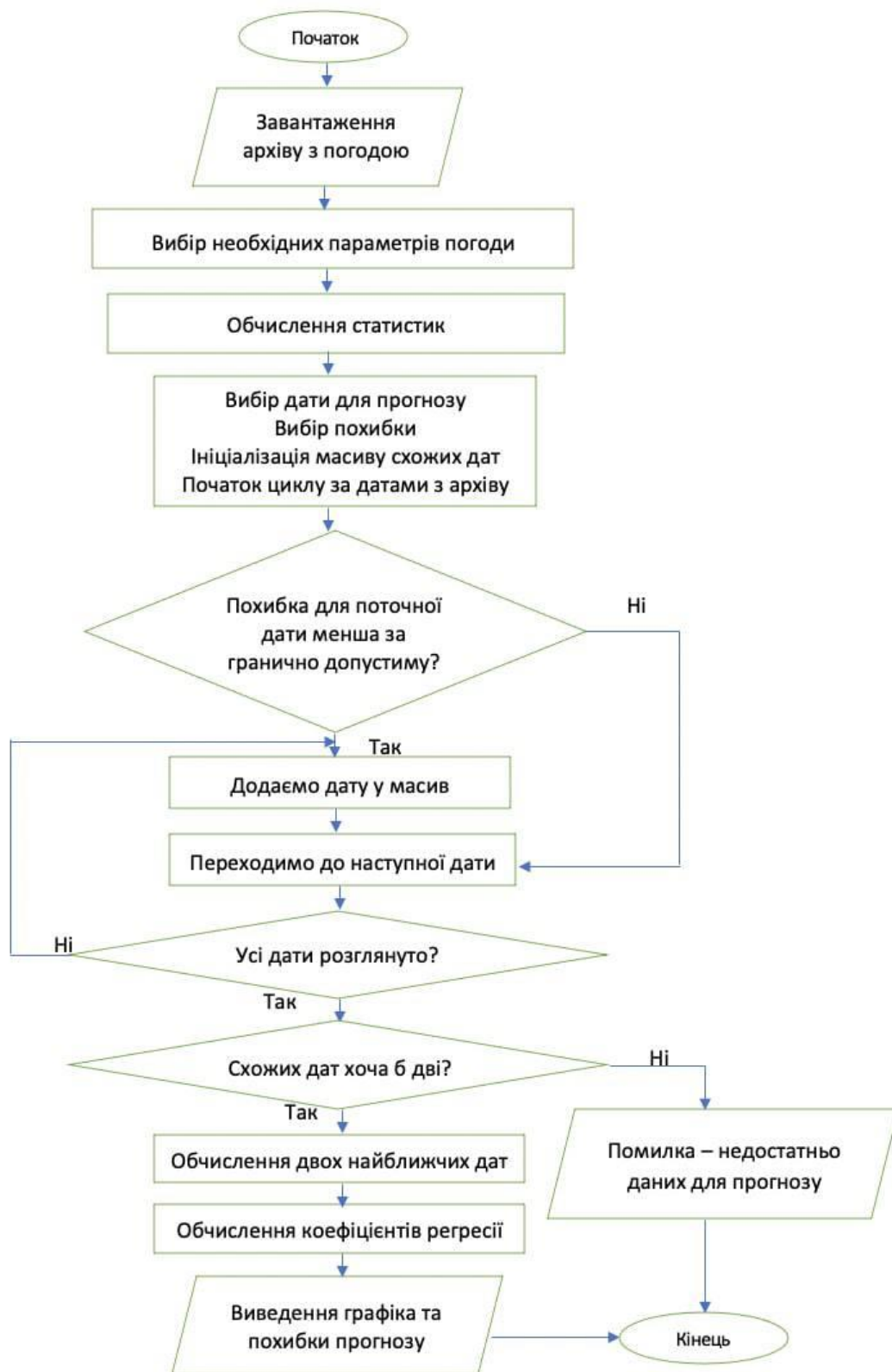


Рис 3.1. Схема алгоритму

Змн.	Арк.	№ докум.	Підпис	Дата

ДП.ДС71.10.000ПЗ

Арк.

43

Розглянемо кожний з етапів детальніше:

1) Завантаження погодинного архіву погоди в Києві.

Одним з небагатьох сайтів, який дозволяє безкоштовно завантажувати детальний архів погоди за досить довгий час (з 2009 року), та який до того ж має для цього зручну бібліотеку **wwow_hist** на мові Python, наведено у [25]. У зв'язку з цим, було прийняте рішення використовувати саме цей сайт і саме цю бібліотеку для завантаження.

2) Вибір необхідних параметрів погоди, від залежить навантаження.

На навантаження впливають температура повітря, атмосферний тиск та відносна вологість повітря.

3) Розрахунок необхідних статистик параметрів за кожну добу, яка міститься в архіві погоди.

Оскільки треба порівнювати дати, кожна з яких містить 24 години, необхідно обрати такі статистики, які б характеризували температуру, тиск та вологість упродовж всієї доби. Таким вимогам краще за все відповідають математичне сподівання та дисперсія.

Таким чином, для кожної дати обчислюватимемо наступні *шість* параметрів: математичне сподівання температури повітря; дисперсія температури повітря; математичне сподівання атмосферного тиску; дисперсія атмосферного тиску; математичне сподівання відносної вологості повітря; дисперсія відносної вологості повітря.

4) Вибір дати, за яку потрібно отримати прогноз навантаження.

Датою, яку було необхідно спрогнозувати погодинне навантаження, стало 4 квітня.

5) Вибір методу пошуку дат, які найбільше схожі (за погодою) на обрану дату.

6) Вибір граничної допустимої розбіжності між датами за обраними статистиками.

Експертним шляхом було встановлено, що максимально припустима похибка вимірювань становить 14%. Отже, якщо хоча б один з шести

					ДП.ДС71.10.000ПЗ	Арк.
						44
Змн.	Арк.	№ докум.	Підпис	Дата		

параметрів цих дат відрізняється більше, ніж на 14%, тоді не можемо вважати ці дати схожими.

7) Обчислення двох найближчих дат.

Для обраної дати перебираємо усі інші дати з усього діапазону дат (з 1 січня 2009 року до 30 квітня 2021 року). Вибираємо з них лише ті, які відповідають критерію, описаному у попередньому пункті. А далі для кожної з тих дат, що залишились, обчислюємо відстань до обраної дати за методом найближчих сусідів, і обираємо ті дві дати, для яких відстань виявилась найменшою.

8) Вибір методу прогнозування навантаження.

Маємо наступну задачу: за кожну з 24 годин маємо температуру, тиск та вологість, а також реальне значення навантаження за цю годину. Необхідно підібрати таку модель, яка буде якісно передбачати значення навантаження за інших значень вхідних параметрів.

Враховуючи малу кількість спостережень (лише 24), було вирішено використати одну з найбільш стійких моделей – модель множинної лінійної регресії, особливості якої було описано у роботі раніше.

9) Обчислення прогнозу погодинного навантаження.

Враховуючи особливості методу лінійної регресії, спостерігаємо параметри моделей, отримані окремо за кожною з «найближчих» дат, можуть суттєво відрізнятись.

Тому, щоб зменшити цей ефект, для обраної дати ми будемо використовувати лінійну комбінацію параметрів цих двох моделей.

$$P_{\text{chosen}} = (p_2 - p_1) + p_2, \quad (3.1)$$

де p_1 – вектор з чотирьох параметрів (коефіцієнти при значеннях температури, вологості та тиску відповідно, та вільний член) для моделі, побудованої за i -ю найближчою датою; $i = 1, 2$; P_{chosen} – вектор параметрів моделі для обраної дати. Знайшовши коефіцієнти P_{chosen} ,

					ДП.ДС71.10.000ПЗ	Арк.
						45
Змн.	Арк.	№ докум.	Підпис	Дата		

можемо обчислити погодинний прогноз навантаження для дати, яку обрали.

10) Побудова графіку навантаження, аналіз результатів.

На цьому етапі ми візуалізуємо реальне та прогнозне навантаження, обчислюємо похибку прогнозу та робимо висновки, чи влаштовують нас результати роботи методів та моделей, та чи не потрібно їх модифікувати.

3.2 Особливості мови програмування

Для програмної реалізації прогнозування навантаження було обрано мову програмування Python.

Python є об'єктно-орієнтованою інтерпретованою мовою програмування [25]. Він надзвичайно простий і містить невелику кількість ключових слів, але, в той же час, дуже гнучкий. Це мова більш високого рівня, ніж Pascal, C++ і, звісно, ніж C. Це досягається в основному за рахунок вбудованих високорівневих структур даних (списки, словники).

Наявність великої кількості модулів, які підключаються до програми і забезпечують додаткові можливості. Такі модулі пишуться як на C, так і на самому Python, і можуть бути розроблені будь-яким досить кваліфікованим програмістом. Як приклад можна навести такі модулі [26]:

- Numerical Python – це розширені математичні можливості, такі як маніпуляції з цілими векторами і матрицями;

- Tkinter – це побудова додатків з використанням графічного інтерфейсу користувача (GUI) на основі широко розповсюдженого на X-Windows Tk-інтерфейсу;

- OpenGL – це використання великої бібліотеки графічного моделювання дво- і тривимірних об'єктів Open Graphics Library фірми Silicon Graphics Inc.

Огляд особливостей:

					<i>ДП.ДС71.10.000ПЗ</i>	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		46

1) Python, на відміну від багатьох мов (Pascal, C ++, Java, і т.д.), не вимагає опису змінних. Вони створюються в місці їх ініціалізації, тобто при першому присвоєнні змінної будь-якого значення. Таким чином, тип змінної визначається типом значення, яке присвоюється. В цьому відношенні Python нагадує Basic.

2) Тип змінної не є незмінним. Будь-яке присвоювання для неї коректно і це призводить лише до того, що типом змінної стає тип нового присвоєння значення.

3) У таких мовах як Pascal, C, C++ організація списків представляла деякі труднощі. Для їх реалізації доводилося добре вивчати принципи роботи з вказівниками і динамічною пам'яттю. І навіть маючи гарну кваліфікацію, програміст, кожен раз заново реалізуючи механізми створення, роботи і знищення списків, міг легко допустити важковловимі помилки. Зважаючи на це, були створені деякі засоби для роботи зі списками. Наприклад, в Delphi Pascal є клас TList, який реалізує списки; для C++ розроблена бібліотека STL (Standard Template Library), що містить такі структури як вектори, списки, множини, словники, стеки і черги. Однак, такі засоби є не у всіх мовах і їх реалізаціях.

4) Python, на відміну від Pascal, C, C++ не підтримує роботу з вказівниками, динамічною пам'яттю і адресну арифметику. У цьому він схожий на Java. Як відомо, вказівники є джерелом важковловимих помилок і робота з ними відноситься більше до низькорівневого програмування. Для забезпечення більшої надійності і простоти вони не були включені в Python.

5) Досить оригінальним є те, як в Python групуються оператори. У Pascal для цього служать операторні дужки begin-end, в C, C ++, Java - фігурні дужки {}, в Basic застосовуються закривають закінчення конструкцій мови (NEXT, WEND, END IF, END SUB).

У мові Python все набагато простіше: виділення блоку операторів здійснюється шляхом зсуву виділеної групи на один або більше пробілів

					ДП.ДС71.10.000ПЗ	Арк.
						47
Змн.	Арк.	№ докум.	Підпис	Дата		

або символів табуляції вправо щодо заголовка конструкції, до якої і буде відноситися даний блок.

Тим самим, хороший стиль запису програм, до якого закликають викладачі мов Pascal, C ++, Java і т.д., тут набувається з самого початку, оскільки інакше просто не вийде.

3.3 Результати роботи програми

1) Завантажуємо погодинні дані рис.3.2:

Retrieving weather data for kiev

Currently retrieving data for kiev: from 2009-01-01 to 2009-01-31
Time elapsed (hh:mm:ss.ms) 0:00:01.054090
Currently retrieving data for kiev: from 2009-02-01 to 2009-02-28
Time elapsed (hh:mm:ss.ms) 0:00:02.118754
Currently retrieving data for kiev: from 2009-03-01 to 2009-03-31
Time elapsed (hh:mm:ss.ms) 0:00:03.279200
Currently retrieving data for kiev: from 2009-04-01 to 2009-04-30
Time elapsed (hh:mm:ss.ms) 0:00:04.243635
Currently retrieving data for kiev: from 2009-05-01 to 2009-05-31
Time elapsed (hh:mm:ss.ms) 0:00:05.452619
Currently retrieving data for kiev: from 2009-06-01 to 2009-06-30
Time elapsed (hh:mm:ss.ms) 0:00:06.614205
Currently retrieving data for kiev: from 2009-07-01 to 2009-07-31
Time elapsed (hh:mm:ss.ms) 0:00:07.710429

executed in 283ms, finished 11:30:41 2021-05-30

◆	date_time ◆	maxtempC ◆	mintempC ◆	totalSnow_cm ◆	sunHour ◆	uvIndex ◆	moon_illumination ◆	moonrise ◆	moonset ◆	sunrise ◆	...	WindGustK
0	2009-01-01 00:00:00	-2	-6	0.2	8.5	1	31	11:14 AM	10:25 PM	08:58 AM	...	
1	2009-01-01 01:00:00	-2	-6	0.2	8.5	1	31	11:14 AM	10:25 PM	08:58 AM	...	
2	2009-01-01 02:00:00	-2	-6	0.2	8.5	1	31	11:14 AM	10:25 PM	08:58 AM	...	
3	2009-01-01 03:00:00	-2	-6	0.2	8.5	1	31	11:14 AM	10:25 PM	08:58 AM	...	
4	2009-01-01 04:00:00	-2	-6	0.2	8.5	1	31	11:14 AM	10:25 PM	08:58 AM	...	

Рис 3.2. Завантаження бази даних погоди

2) Обираємо необхідні параметри – температуру повітря та вологість (за результатами дослідження, прогноз без урахування атмосферного тиску виявився точнішим, якщо тиск має за високу дисперсію) на рис. 3.3.

executed in 130ms, finished 14:06:29 2021-06-04

⚡	date ⚡	hour ⚡	tempC ⚡	humidity ⚡
0	2009-01-01	0	-6	97
1	2009-01-01	1	-6	97
2	2009-01-01	2	-6	97
3	2009-01-01	3	-6	97
4	2009-01-01	4	-6	97

Рис 3.3. Обирані параметри

3) Розраховуємо математичні сподівання та дисперсії цих параметрів за кожну добу (рис. 3.4):

⚡	date ⚡	tempC_mean ⚡	humidity_mean ⚡	tempC_variance ⚡	humidity_variance ⚡	day_of_week ⚡
0	2009-01-01	-3.875000	96.041667	2.276042	2.289931	3
1	2009-01-02	-6.875000	95.333333	4.442708	3.055556	4
2	2009-01-03	-7.583333	95.500000	2.909722	0.500000	5
3	2009-01-04	-5.666667	95.125000	1.138889	3.276042	6
4	2009-01-05	-8.791667	96.666667	7.914931	2.138889	0

Рис 3.4. дисперсія і математичне сподівання

4) Фільтруємо дати з архіву за критерієм схожості (для математичного сподівання температури замінюємо відносну похибку на абсолютну, через від'ємне значення температури).

5) Таким чином, залишилось лише вісім дат, з яких потрібно знайти дві найближчих (рис. 3.5). Зробимо це за допомогою методу найближчих сусідів:

('2017-04-05', '2010-09-21')

Рис 3.5. Знайдені найближчі дві дати

7) Готуємо дані для навчання лінійної регресії. Для цього спочатку завантажуюмо реальні значення погодинного навантаження (рис. 3.6):

executed in 15ms, finished 15:42:28 2021-06-04		
⚡	hour ⚡	power ⚡
0	0	0.03120
1	1	0.02912
2	2	0.02496
3	3	0.02080
4	4	0.02080
5	5	0.03328
6	6	0.07488
7	7	0.08320
8	8	0.12480
9	9	0.13520
10	10	0.13728
11	11	0.14560
12	12	0.13520
13	13	0.13312
14	14	0.11440
15	15	0.10400
16	16	0.16640
17	17	0.17056
18	18	0.16848
19	19	0.18720
20	20	0.17680
21	21	0.16016
22	22	0.09360
23	23	0.05824

Рис 3.6. Реальне значення навантаження

А тепер отримуємо повні дані для обох найближчих дат. Ці дані зображені на рис. 3.7.

⚡	date ⚡	hour ⚡	tempC ⚡	humidity ⚡	power ⚡	⚡	date ⚡	hour ⚡	tempC ⚡	humidity ⚡	power ⚡
0	2017-04-05	0	10	97	0.03120	0	2010-09-21	0	10	99	0.03120
1	2017-04-05	1	9	97	0.02912	1	2010-09-21	1	10	99	0.02912
2	2017-04-05	2	9	98	0.02496	2	2010-09-21	2	9	99	0.02496
3	2017-04-05	3	9	98	0.02080	3	2010-09-21	3	9	99	0.02080
4	2017-04-05	4	10	94	0.02080	4	2010-09-21	4	10	96	0.02080
5	2017-04-05	5	10	89	0.03328	5	2010-09-21	5	11	93	0.03328
6	2017-04-05	6	11	85	0.07488	6	2010-09-21	6	12	90	0.07488
7	2017-04-05	7	12	81	0.08320	7	2010-09-21	7	13	81	0.08320
8	2017-04-05	8	13	78	0.12480	8	2010-09-21	8	15	72	0.12480
9	2017-04-05	9	13	75	0.13520	9	2010-09-21	9	16	63	0.13520
10	2017-04-05	10	15	66	0.13728	10	2010-09-21	10	17	58	0.13728
11	2017-04-05	11	16	57	0.14560	11	2010-09-21	11	18	53	0.14560
12	2017-04-05	12	18	49	0.13520	12	2010-09-21	12	18	48	0.13520
13	2017-04-05	13	17	49	0.13312	13	2010-09-21	13	18	53	0.13312
14	2017-04-05	14	17	49	0.11440	14	2010-09-21	14	17	59	0.11440
15	2017-04-05	15	17	50	0.10400	15	2010-09-21	15	16	64	0.10400
16	2017-04-05	16	15	55	0.16640	16	2010-09-21	16	14	69	0.16640
17	2017-04-05	17	14	61	0.17056	17	2010-09-21	17	13	74	0.17056
18	2017-04-05	18	12	67	0.16848	18	2010-09-21	18	11	80	0.16848
19	2017-04-05	19	11	70	0.18720	19	2010-09-21	19	11	81	0.18720
20	2017-04-05	20	11	74	0.17680	20	2010-09-21	20	11	82	0.17680
21	2017-04-05	21	10	77	0.16016	21	2010-09-21	21	11	83	0.16016
22	2017-04-05	22	10	79	0.09360	22	2010-09-21	22	10	83	0.09360
23	2017-04-05	23	9	81	0.05824	23	2010-09-21	23	10	83	0.05824

Рис 3.7. дані обох найближчих дат

8) Отримуємо комбіновану модель та навчаємо її.

Передбачення вектору погодинних навантажень за допомогою навченої моделі зображено на рис. 3.8.

predicted
executed in 45ms, finished 15:42:37 2021-06-04
[0.025950377911717837, 0.025950377911717837, 0.0563402329289977, 0.0563402329289977, 0.050119894514909724, 0.051956061635219264, 0.08418208377280845, 0.07174140694463249, 0.09774709066813414, 0.08530641383995818, 0.13364544704634218, 0.1515946252354462, 0.16954380342455022, 0.16148729789015293, 0.1838206473730355, 0.1757641418386382, 0.18820481866681404, 0.16219913494331273, 0.17463981177148857, 0.15852680070269398, 0.17280364465117914, 0.15669063358238455, 0.11640810591039785, 0.08418208377280845]

Рис 3.8. Спрогнозоване значення навантаження

9) Представляємо передбачені та реальні значення погодинного навантаження у вигляді графіку (рис. 3.9).

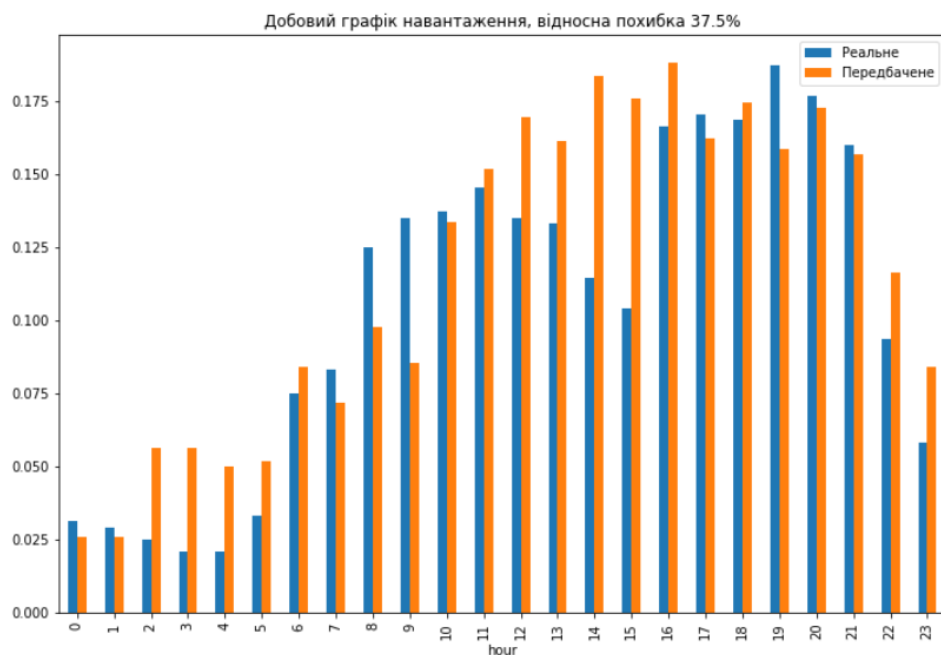


Рис 3.9. Передбачені та реальні значення погодинного навантаження

ВИСНОВКИ

Було досліджено та проаналізовано методи прогнозування навантаження в системах розподілених джерел енергії, та обрано регресивний метод прогнозування. Головна перевага даного методу - це простота розрахунків, прозорість моделювання, одноманітність аналізу. В ході дослідження була створена база даних навколишнього середовища (масив даних температур, тиску, вологості за певний період), для побудови регресивних рівнянь на основі яких в подальшому розрахована прогнозна модель. Пошук в базі даних навколишнього середовища відбувалася за допомогою обраних пошукових параметрів (математичного очікування та дисперсії). Під час пошуку найбільш схожих даних, було виявлено, що різниця між вхідними і знайденими даними не повинна перевищувати 14%, для адекватності прогнозної моделі. В результаті була створена прогнозна модель на основі регресивних рівнянь схожих за параметрами днів. Провівши аналіз прогнозованого навантаження та порівнявши його з реальними даними, отримали похибку в 41%.

					<i>ДП.ДС71.10.000ПЗ</i>	Арк.
						52
Змн.	Арк.	№ докум.	Підпис	Дата		

ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Евланов, Л. Г. Экспертные оценки в управлении / Л. Г. Евланов, В. А. Кутузов. - М.: Экономика, 1978. - 134 с.
2. Кудрин, Б. И. Методика обеспечения почасового прогнозирования электропотребления предприятий с учетом погодных факторов [Текст] / Б. И. Кудрин, А. В. Мозгалин // Вестник МЭИ. - 2007. - № 2. - С. 105 - 108.
3. Моделі і методи прогнозування [Електронний ресурс] – Режим доступу:http://infolibrary.com/content/2122_Metodi_i_modeli_prognozyvanny_a.htm// (дата звернення: 08.04.2021)
4. Особливості короткострокового прогнозування електричного навантаження енергосистеми із суттєвою складовою промислового електроспоживання / П. О. Черненко, О. В. Мартинюк, В. О. Мірошник // Праці Інституту електродинаміки Національної академії наук України. - 2016. - Вип. 43. - С. 24-31.
5. Jingfei Yang M. Sc. Power System Short-term Load Forecasting: Thesis for Ph.d degree. Germany, Darmstadt, Elektrotechnik und Informationstechnik der Technischen Universitat, 2006. 139 p.
6. Льюїс, К. Д. Методи прогнозування економічних показників / К. Д. Льюїс; перекл. з англ.: Е. З. Демиденко. - М.: Фінанси и статистика, 1986. - 132 с
7. Андерсон, Т. Статистичний аналіз часових рядів / Т. Андерсон; Перекл. з англ. - М.: Мир, 1976. - 760 с.
8. Методи прогнозування. [Електронний ресурс] – Режим доступу: <https://library.if.ua/book/72/5247.html> //(дата звернення:18.04.21)
9. Prajakta S.K. Time series Forecasting using Holt-Winters Exponential Smoothing // Kanwal Rekhi School of Information Technology Journal 2004. 13p.

					ДП.ДС71.10.000ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		53

10. Pradhan R.P., Kumar R. Forecasting Exchange Rate in India: An Application of Artificial Neural Network Model // Journal of Mathematics Research. 2010, Vol. 2, No. 4. P. 111 - 117.

11. Владимирова Л. П., Прогнозирование и планирование в условиях рынка: Учебное пособие. 2-е изд., перераб. и доп. – М.: Дашков и К, 2008. – 307 с

12. Белов А.А., Кропотов Ю.А., Проскуряков А.Ю. Автоматизированный анализ и обработка временных рядов данных о загрязняющих выбросах в системе экологического контроля // Информационные системы и технологии, 2010. – №6(62). – С. 28– 35.

13. Щербаков М. В., Щербакова Н. Л., Козлов И. П., Камаев В. А. Коннективистские модели идентификации динамики систем на коротких интервалах наблюдения с заданным множеством классов поведения // Известия ЮФУ. Технические науки. – 2012. – №1. – С. 83-91.

14. Курбацкий, В. Г. Прогнозирование электрической нагрузки с использованием искусственных нейронных сетей [Текст] / В. Г. Курбацкий, Н. В. Томин // Электрика. - 2006. - №- С. 26 - 32.

15. Кацко, И.А. Практикум по анализу данных на компьютере. Под ред. Г.В.Гореловой. – М.: КолосС, 2009. – 278с

16. Е.М. Четыркин. Статистические методы прогнозирования. – М.: Статистика, 1977. – 199 с.

17. Меламед М.А. Современные методы анализа и прогнози-рования режимов электропотребления в электроэнергетических системах // Итоги науки и техники. – Сер. Энергетические системы и их автоматизация. – 1988. – Т.4. – С. 4–11.

18. Комашинский В.И., Смирнов Д.А. Нейронные сети и их применение в системах управления и связи: Учебное пособие. – М.: 2012. – 94 с.

19. Басовский Л. Е. Прогнозирование и планирование в условиях рынка. Учеб. пособие.- М. : ИНФРА-М, 2011

					<i>ДП.ДС71.10.000ПЗ</i>	Арк.
						54
Змн.	Арк.	№ докум.	Підпис	Дата		

20. Козьева А., Вертакова Ю. В., Бабич Т. Н. Прогнозирование и планирование в условиях рынка. Учеб. пособие для вузов. – М. : ИНФРА-М, 2012.

21. Саркисян С. А. Теория прогнозирования и принятия решений / С.А. Саркисян, В.И. Каспин, В.А. Лисичкин, Э.С. Минаев, Г.С. Пасечник. – М.: Высшая школа, 1977. – 351 с.

22. J. Yamnenko, T. Tereshchenko, L. Klepach та D. Palii, «Forecasting of electricity consumption in SmartGrid.,» в Proceedings of the International Conference on Modern Electrical and Energy Systems, Kremenchuk, 2017.

23. S. Soliman, Electrical load forecasting: modeling and model construction. Burlington Butterworth-Heinemann, 2016.

24. [17] В. Доманов та А. Билалова, «Анализ прогнозирования энергопотребления с различными информационными базами.,» Известия Самарского научного центра Российской академии наук., т. 16, pp. 530-535, 2014.

25. Особливості Python. [Електронний ресурс] – Режим доступу: <https://www.geeksforgeeks.org/python-features/> (дата звернення: 4.06.21)

26. Python (мова програмування). [Електронний ресурс] – Режим доступу: [https://en.wikipedia.org/wiki/Python_\(programming_language\)](https://en.wikipedia.org/wiki/Python_(programming_language)) (дата звернення: 4.06.21)

					ДП.ДС71.10.000ПЗ	Арк.
						55
Змн.	Арк.	№ докум.	Підпис	Дата		

ЛІСТИНГ ПРОГРАМИ

Програмний код для прогнозування навантаження

```
from wwo_hist import retrieve_hist_data

frequency = 1
start_date = '01-JAN-2009'
end_date = '30-APR-2021'
api_key = '32bbb7f5a520495e874170438212705'
location_list = ['kiev']

hist_weather_data = retrieve_hist_data(api_key,
                                       location_list,
                                       start_date,
                                       end_date,
                                       frequency,
                                       location_label = False,
                                       export_csv = True,
                                       store_df = True)

import pandas as pd
kyiv_weather_all = pd.read_csv('kiev.csv')
kyiv_weather_all.head()

kyiv_weather_all['date'] = kyiv_weather_all.date_time.apply(lambda x:
x[:10])
kyiv_weather_all['hour'] =kyiv_weather_all.date_time.apply(lambda x:
int(x[11:13]))

needed_fields = ['tempC', 'humidity']

kyiv_needed_data = kyiv_weather_all[['date', 'hour'] + needed_fields]
kyiv_needed_data.head()

import numpy as np
from datetime import datetime
```

```

        kyiv_statistics_mean = kyiv_needed_data.drop('hour',
axis=1).groupby('date').apply(np.mean).reset_index()
        kyiv_statistics_variance = kyiv_needed_data.drop('hour',
axis=1).groupby('date').apply(np.var).reset_index()

        kyiv_statistics_mean.columns = ['date'] + [x + '_mean' for x in
needed_fields]
        kyiv_statistics_variance.columns = ['date'] + [x + '_variance' for x in
needed_fields]

        kyiv_statistics_all =
kyiv_statistics_mean.merge(kyiv_statistics_variance)
        kyiv_statistics_all['day_of_week'] =
kyiv_statistics_all['date'].apply(lambda x: datetime.strptime(x, '%Y-%m-
%d').weekday())

        kyiv_statistics_all.head()

        chosen_date = '2012-04-20'
        chosen_stats = kyiv_statistics_all[kyiv_statistics_all['date'] ==
chosen_date]
        cur_vals = chosen_stats.values[0]
        chosen_stats

        data_list = []

        for new_vals in kyiv_statistics_all.values:
            if new_vals[0] == cur_vals[0]:
                continue
            can = True
            if abs(new_vals[1] - cur_vals[1]) > 2:
                continue
            for idx in range(2, len(new_vals)-1):
                if abs(new_vals[idx] - cur_vals[idx]) /
max(max(abs(new_vals[idx]), abs(cur_vals[idx])), 0.01) > 0.12:
                    can=False
                    break
            if can:
                data_list.append(new_vals)
        data_list

        similarity_scores = []

```

```

for new_vals in data_list:
    score = 0
    for idx in range(1, len(new_vals)-1):
        score += (new_vals[idx] - cur_vals[idx]) * (new_vals[idx] -
cur_vals[idx])
    similarity_scores.append((score, new_vals[0]))

best_pair = sorted(similarity_scores)[:2]

first_date = best_pair[0][1]
second_date = best_pair[1][1]

first_date, second_date

power_data = pd.read_excel(u'Навантаження.xlsx')
power_data.columns = ['hour', 'power']
power_data

first_date_values      =      kyiv_needed_data[kyiv_needed_data.date      ==
first_date]
second_date_values     =      kyiv_needed_data[kyiv_needed_data.date     ==
second_date]

first_date_total_values = first_date_values.merge(power_data)
second_date_total_values = second_date_values.merge(power_data)

display(first_date_total_values)
display(second_date_total_values)

Y1 = first_date_total_values['power'].values
X1 = first_date_total_values[needed_fields].values

Y2 = second_date_total_values['power'].values
X2 = second_date_total_values[needed_fields].values

from sklearn.linear_model import LinearRegression
reg1 = LinearRegression().fit(X1, Y1)
reg2 = LinearRegression().fit(X2, Y2)

reg_final_coef = (reg2.coef_ - reg1.coef_) + reg2.coef_
reg_final_intercept = (reg2.intercept_ - reg1.intercept_) +
reg2.intercept_

```



```

    kyiv_data_for_chosen_date = kyiv_needed_data[kyiv_needed_data.date ==
chosen_date]
    predicted = []

    for hour in kyiv_data_for_chosen_date.hour:
        hour_data = kyiv_data_for_chosen_date[kyiv_data_for_chosen_date.hour
== hour].values[0]
        predicted.append(hour_data[2:].dot(reg_final_coef) +
reg_final_intercept)

    predicted

    from matplotlib import pyplot as plt

    plt.figure(figsize=(12, 8))
    plt.title(u'Порівняння реального навантаження з передбаченим')
    plt.xlabel(u'Година')
    plt.ylabel(u'Навантаження')
    plt.ylim([0, 0.3])
    plt.plot(range(24), power_data.power)
    plt.plot(range(24), predicted)
    plt.legend([u'Реальне', u'Передбачене'])
    plt.show()

```

S U M M A R Y

Microgrid load of forecasting

The diploma project of first educational level "Bachelor" by specialty 171 Electronics, specialization electronic components and systems Rostislav Marchenko. National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute». Faculty of Electronics, Department of Electronic Devices and Systems. Academic group DS-71. - Kyiv: Igor Sikorsky Kyiv Polytechnic Institute, 2021. - 5 p., Ill. tables

Keywords: load of forecasting, regression.

Summary of the project: 5 pages

S U M M A R Y

At present, there is a question of forecasting the load of local facilities, according to the environment. This is relevant for companies involved in electricity supply. These companies are required to report the planned hourly electricity consumption schedule for the day ahead. Deviation from the planned schedule results in additional financial losses for the purchase or sale of excess electricity at a low price. Also, electricity consumption forecasting is relevant for industrial facilities and private homes. Currently, forecasting is most commonly used in Microgrid, a localized grouping of distributed energy sources such as solar, wind, hydropower, and storage and backup and load management tools. This system can work as a separate entity, or its users can connect to a larger network as needed. Micro-grid growth is expected in developed countries that increase the resilience of electricity grids, as well as in developing countries in areas that currently do not have access to electricity using alternative renewable energy sources such as hydropower, distributed solar photovoltaics. It is assumed that these systems

replace what would otherwise be a continuation of the dirty power grid or the subsequent use of off-grid oil or diesel generators.

Therefore, load forecasting will always be relevant, especially with the proliferation of systems such as Microgrid.

Review of forecasting methods

The forecasting method is a method of studying an object with a certain sequence of actions that must be performed to obtain a forecasting model.

Forecasting methods are divided into intuitive, formalized, combined.

Intuitive methods are based on the information of collective expert assessments, this method is useful in the process of creating a predictive load model. The accuracy of the forecast depends on the qualifications of the expert, but due to the high degree of dependence on the quality of the specialist's forecast, we can conclude that intuitive methods are not suitable for creating models of load forecasting. Also intuitive methods useful in creating a predictive model based on formalized methods.

Formalized methods are used with a sufficient amount of collected information, namely about the current and past state of the object of study.

Combined methods are methods with a mixed information base, in which both expert and factual information is used as the primary one.

A time series is a set of data with a certain sequence at different points in time. Time series are divided into statistical and structural.

The statistical method helps to find the relationship between the load and the external factor that influenced the data in the past and so we can obtain the equation of the relationship between the load and the impact factor. Therefore, statistical methods can predict the load of any day.

The most common methods of forecasting time series:

- correlation and regression analysis;
- predictive extrapolation;
- adaptive forecasting methods;

- autoregressive analysis;
- exponential smoothing;
- classification and regression trees;

To build a model of multiple regression, it is necessary to determine the regression coefficients using information about the object of study and the factors that affect it. It follows that correlation and regression methods are able to take into account various parameters that affect the object of forecast.

Estimation of load forecasting

The adequacy of the model is the degree of its compliance with the real object of study. The model is considered adequate if the systematic component (trend) is defined correctly.

The reliability of the model is characterized by the level of proximity of the calculated levels of the series to the actual values. It should be noted that the process of assessing the quality of the forecast begins with checking the quality of the original data, the correctness of the theoretical methods of forecasting, which form the basis of the forecast model.

The quality of the source information depends on the accuracy of measurements, the range of data studied, the chosen research method, the accuracy of the results.

Criteria for assessing the quality of the forecast are used to obtain characteristics such as: reliability, accuracy, reliability, forecast error.

The accuracy of the prediction is assessed by comparing the actual and predicted values. The accuracy of the forecast is through such quantities as:

- root mean square error of the forecast,
- average absolute error of the forecast in percent,
- average absolute forecast error.

Considering the methods of load prediction in MicroGrid, the most suitable is the method of regression analysis, because the behavior of the load curve is described by a function of time, so we choose models of time series.

The main task of the forecast model is to predict the daily value of the load, based on weather conditions and preliminary observation (dependence of the load on the external factor). Input parameters of the forecast model: daily values of hourly load, parameters of external conditions.

To find the relationship between the load and the environment, factors such as temperature, pressure, humidity are selected.

Let's check whether there is a statistical relationship between the load parameters and the environmental parameters, calculate the regression coefficients.

Therefore, based on the results of the calculated correlation coefficients, we conclude that the selected study data are statistically dependent.

This dependence allows you to create a predictive model based on the results of previous observations.

For the obtained similar days we calculate regressive equations.

We take the first similar day found and calculate the regression coefficients and check them for significance to obtain a regression equation.

Calculating the regression coefficients obtained by the equation, we check the statistical significance of the equation.

Based on the results of checks of the Fisher criterion and the coefficient of determination, we conclude about the statistical significance of the equation and found that the parameters of the model are statistically significant.

In the same way, the significance and multiple regression equation of the second selected day were calculated and verified.

Having the regression equations of two similar days, we calculate the first finite difference between the regression coefficients, subtracting the first from the second equation.

Using the obtained regression equation, calculate the predicted load. To do this, substitute in the equation environmental data (X1-temperature, X2-pressure, X3-humidity). And we get the projected daily load.

Having the data of real and projected load, we calculate the error of the forecast model according to the formula of the average absolute error MAPE.

We plot the actual and projected load to see graphically the point difference of the forecast.

Conclusions: load forecasting is especially important in autonomous operation, as renewable power sources are characterized by high instability of the level of generated power, which depends on climatic conditions. Therefore, the urgent task for MicroGrid is to predict the load under different environmental conditions. Among the considered methods for load forecasting in MicroGrid systems, the most suitable is the method of regression analysis, as it is possible to create an automated forecasting system. In addition, the behavior of the energy consumption curve is described as a function of time, so we choose a group of models of time intervals.