

ПІДВИЩЕННЯ ТОЧНОСТІ ПРОГНОЗУ ВИРОБНИЦТВА ЕЛЕКТРОЕНЕРГІЇ ФОТОЕЛЕКТРИЧНИМИ СТАНЦІЯМИ НА ОСНОВІ МЕТОДУ RANDOM FOREST

¹Вінницький національний технічний університет

Досліджено методи, що використовуються для прогнозування виробництва енергії фотоелектричними станціями (ФЕС), а також шляхи підвищення точності прогнозу для оптимізації структури балансу електроенергії в енергосистемі. Дослідження спрямоване на виявлення ефективних підходів та алгоритмів прогнозування, оцінювання їхньої точності та надійності. За результатами дослідження запропоновано комбінацію методів регресійного аналізу та машинного навчання, що забезпечує прийнятну точність прогнозу для планування резервів потужності у енергосистемі. За основу взято метод Random Forest, оскільки він забезпечує адаптивність до особливостей енергогенерування ФЕС у різних регіонах України та в різні періоди року. Для підвищення ефективності машинного навчання запропоновано алгоритм попередньої фільтрації наборів даних з використанням методів авторегресії та ковзного середнього. Це дозволяє краще підготувати вхідні дані для подальшого використання у прогнозуванні, забезпечуючи згладжування тимчасових рядів і усунення випадкових коливань, які можуть негативно впливати на точність прогнозу. Застосування методів попередньої фільтрації дозволяє виділити основні закономірності в даних, що також підвищує точність моделей машинного навчання. Метод Random Forest вибрано не випадково: він добре підходить для завдань прогнозування, де необхідно враховувати багато різних факторів, які можуть по-різному впливати на результати в залежності від часу та регіону. Це особливо важливо у випадку прогнозування генерації енергії ФЕС, де на виробництво значно впливають такі змінні, як хмарність, температура, сезонність тощо. Використання Random Forest дозволяє враховувати нелінійні залежності та взаємодію між факторами, що сприяє підвищенню точності прогнозів. Застосування Random Forest як основного алгоритму машинного навчання зумовлене його гнучкістю та здатністю адаптуватися до різних умов, що дозволяє ефективно враховувати особливості генерації енергії в різних регіонах країни та у різні сезони року.

Ключові слова: фотоелектричні станції, генерування, прогнозування, регресійний аналіз, машинне навчання, аналіз даних.

Вступ

Сучасні енергосистеми стикаються з викликами, пов'язаними зі збільшенням частки відновлюваних джерел енергії (ВДЕ), зокрема фотоелектричних станцій (ФЕС), у структурі балансу електроенергії. Ці джерела є досить ефективними з екологічної точки зору, проте через значну залежність генерування від погодних умов, вони ускладнюють прогнозування та підтримку балансу електроенергії в електроенергетичній системі (ЕЕС), планування резервів та ведення режиму. Особливо це проявляється в умовах обмеженості ресурсів для регулювання частоти і потужності, що характерно для енергосистеми України. Для вирішення проблеми використовують примусові обмеження генерування ВДЕ, що особливо негативно відбивається на рентабельності ФЕС. Більше того, виникає необхідність залучати аварійну допомогу від ENTSO-E, що знижує прибутковість енергоринку України.

Одним зі шляхів вирішення проблеми є підвищення ефективності прогнозування графіків генерування ФЕС [1] разом із застосуванням засобів внутрішньодобового перерозподілу енергії. Якісний прогноз дає змогу передбачити потенційні коливання у виробництві електроенергії від ВДЕ і вжити відповідних заходів для формування резервів потужності в ЕЕС у певні періоди доби, вико-

ристовуючи наявні, обмежені ресурси. Таким чином, якісний прогноз виробництва енергії з ВДЕ безпосередньо сприяє підвищенню їхньої економічної ефективності, а також забезпеченню стійкості й надійності функціонування ЕЕС та її операційній безпеці.

Останнім часом для прогнозування широко застосовуються методи машинного навчання та штучні нейронні мережі. Ці алгоритми базуються на аналізі великих обсягів даних та виявленні складних зв'язків між вихідними даними [2] (наприклад, погодними умовами та генеруванням електроенергії ФЕС). Процес оцінювання графіків генерування ФЕС має враховувати низку характеристик самої станції, таких як встановлена потужність, схемні рішення, стан та орієнтування сонячних панелей, наявність та тип трекерів, а також наявність засобів регулювання генерування залежно від потреб енергосистеми, засобів накопичення електроенергії тощо. Важливо зазначити, що прогнозування генерування ФЕС є складним завданням. Воно пов'язане з необхідністю прогнозування погодних умов в зоні розташування сонячних панелей. А ці умови мають значну стохастичну (складно-прогнозовану) складову. Раптові зміни погодних умов, особливо хмарності, призводять до значних коливань генерування енергії ФЕС (від 20 % до 90 %). Через що точність прогнозів істотно залежить від доступності та актуальності даних про впливові параметри навколишнього середовища, від точності реєстрації графіків генерування, а також від ефективності вибраних алгоритмів прогнозування.

Прогрес у галузі прогнозування генерування ФЕС дає змогу енергетичним компаніям та операторам систем передачі (ОСП) підвищити ефективність формування оперативних резервів енергії для балансування. Це зменшує негативний вплив негарантованого генерування від ВДЕ на операційну безпеку ЕЕС. Таким чином, якісне прогнозування графіків видачі енергії ФЕС відіграє важливу роль у досягненні сталого розвитку енергосистеми. Зокрема, це дає змогу збільшити допустимі обсяги заміщення традиційного генерування відновлюваними джерелами, забезпечує передумови переходу до чистої та стійкої енергетики, мінімізуючи негативний вплив на довкілля та оптимізуючи процеси в енергосистемі.

Метою роботи є аналіз методів прогнозування генерування фотоелектричних станцій, виявлення їхніх особливостей та розроблення заходів з підвищення точності прогнозування графіків роботи ФЕС на основі методів машинного навчання.

Аналіз досліджень і публікацій

Аналіз публікацій підтверджує, що статистичні методи свого часу виявилися дуже корисними для прогнозування генерування ФЕС. Метод авторегресії (*Auto Regression — AR*) є статистичним методом аналізу часових рядів та прогнозування, який базується на ідеї, що поточне значення змінної залежить від її попередніх значень у часі. Цей метод використовує ретроспективні дані для побудови залежностей і визначення параметрів, які характеризують зміну процесу дослідження в часі [3]. Ідентифікація *AR*-моделі включає в себе визначення її порядку p та оцінювання коефіцієнтів. Для цього застосовуються статистичні методи, зокрема такі як критерії Акаїке чи Байєса [2]. Метод авторегресії широко використовується в економетриці, фінансах та інших галузях для прогнозування часових рядів і аналізу їхньої динаміки [2].

Іншим класичним підходом до аналізу числових рядів є метод ковзного середнього (*Moving Average — MA*). Він дає змогу враховувати вплив випадкового шуму та випадкових факторів на процес генерування ФЕС. Об'єднання цих двох методів у метод *ARMA* дає змогу аналізувати як довгострокові, так і короткострокові часові залежності. Прогнозування обсягів генерування ФЕС за допомогою статистичних методів відрізняється високою точністю та надійністю, особливо на короткостроковому та середньостроковому горизонті [4]. Дослідження підтверджують цю точність, зокрема, для методу *ARIMA (Autoregressive Integrated Moving Average)*. А варіація методу *ARIMA-ANN (ARIMA with Artificial Neural Network)* виявила високу ефективність й у короткостроковому прогнозуванні обсягів генерування ФЕС [4].

Методи машинного навчання, зокрема нейронні мережі, набули популярності у прогнозуванні генерування ФЕС. Вони дають змогу виявляти складні нелінійні залежності між різними вихідними змінними, такими як сонячна радіація, температура, вологість та час доби. Методи машинного навчання, зокрема метод опорних векторів (*SVM*) та ансамблеві методи, також успішно використовуються для короткострокового прогнозування генерування ФЕС [5], [6].

Часові нейронні мережі (*Time Series Neural Networks*) [7] є спеціальним типом нейронних мереж, призначених для роботи з часовими рядами даних. Вони можуть автоматично виявляти скла-

дні закономірності та нелінійні зв'язки між різними змінними. Це дозволяє їм прогнозувати майбутні значення з високою точністю, навіть у складних умовах. Часові нейронні мережі можуть бути навчені за допомогою різних алгоритмів, таких як зворотне поширення помилки *BP* (*Backpropagation*) [8] або алгоритми генетичного навчання [9], щоб оптимізувати їхні параметри для кожної конкретної задачі прогнозування.

Для підготовки та первинної обробки часових рядів змінних, що використовуються для прогнозування, на практиці ефективно застосовують експоненційне згладжування (*Exponential Smoothing*) [10]. Основна ідея полягає в тому, щоб надати більшої ваги останнім спостереженням, а нестабільним чи випадковим варіаціям — меншої. Це дає змогу підвищити якість навчальної вибірки для прогнозування в ситуаціях, коли тренд чи сезонні коливання є не дуже виразними. В залежності від характеру значень змінних у наборі даних застосовуються різні підтипи експоненційного згладжування: просте, подвійне, або потрійне.

Важливим аспектом прогнозування генерування ФЕС є інтегрування погодних моделей, оскільки погодні умови безпосередньо впливають на виробництво електричної енергії. Врахування погодних чинників, таких як хмарність, вітрові умови, опади, може значно покращити точність прогнозування генерування ФЕС порівняно з регресійним аналізом часових рядів [8]. Недавні дослідження свідчать, що поєднання інформації про погоду зі статистичними моделями може значно підвищити точність прогнозування генерування ФЕС [11].

У підсумку зауважимо, що статистичні методи, які зазвичай ґрунтуються на припущенні про лінійність процесів у часі, мають ускладнення з врахуванням нелінійних зв'язків між окремими факторами впливу. Тобто такі методи не відповідають поставленому завданню.

Як показує практика, щоб підвищити точність прогнозування погодинного графіка генерування ФЕС, доцільно застосовувати методи на основі штучного інтелекту. Так для прогнозування потужностей ФЕС широко використовують методи, засновані на стратегіях машинного навчання, зокрема *Random Forest* [12], [13]. Порівняно з методами, що ґрунтуються на статистиці, підходи машинного навчання є гнучкішими та інтелектуальними, тобто можуть виявляти складні нелінійні взаємозв'язки на основі великих обсягів даних. Проте традиційні моделі машинного навчання мають відносно обмежену адаптивність до змін у розподілі даних і нелінійних взаємозв'язків [14]. Тобто під час розв'язання складних нелінійних завдань зазвичай доводиться вручну вводити нелінійні ознаки, або відповідним чином коригувати дані датасетів. Виходом є застосування методів глибокого навчання, здатних до наскрізного навчання й автоматичного вилучення нехарактерних ознак. Їх часто застосовують у галузі прогнозування складних процесів, зокрема викидів вуглецевих газів [15]. Проте такий підхід пов'язаний з необхідністю опрацювання значних обсягів даних та залучення великих обчислювальних ресурсів. Для задачі прогнозування погодинних графіків генерування сотень окремих ФЕС, що розміщені у різних регіонах країни, це може бути істотним обмеженням.

Результати дослідження

Підвищення точності прогнозування графіків виробництва електроенергії ФЕС є актуальним питанням для України та інших країн. Враховуючи посилення вимог до точності прогнозування генерування ФЕС, а також зміни кліматичних умов, особливо протягом останніх років, відомі методи прогнозування потребують постійного аналізу та удосконалення. Як базовий варіант для формування графіків генерування ФЕС на добу наперед розглянемо застосування технології машинного навчання на основі методу *Random Forest* [12]. *Random Forest* — це ансамблевий метод машинного навчання, який використовується для класифікації даних та формування моделі регресії. Основна ідея полягає у побудові численних дерев прийняття рішень під час тренування моделі й продукування моди для класів або осереднення регресії побудованих дерев. Перевагами методу є [13]:

- здатність ефективно обробляти дані з великою кількістю ознак і класів;
- нечутливість до масштабування значень ознак;
- можливість обробляти неперервні й дискретні ознаки;
- внутрішня система оцінювання здатності побудованої моделі до узагальнення;
- можливість організації паралельних обчислень з багатьма потоками.

Random Forest використовує багато різних дерев прийняття рішень для досягнення точніших та стійких прогнозів.

Основні етапи застосування *Random Forest* для формування прогнозової моделі:

1. *Формування ансамблю.* Алгоритм створює випадковий ансамбль дерев прийняття рішень. Кожне дерево формується на основі випадкової підмножини навчальних даних та на основі випадково вибраних факторів для розгляду кожного рішення.

2. *Прогнозування кожним деревом.* Кожне дерево в ансамблі прогнозує результат на основі своєї власної підмножини даних. Для отримання узагальненого прогнозу прогнози цих дерев об'єднуються.

3. *Оцінка і ваговий внесок.* Кожне дерево отримує вагу з урахуванням точності його прогнозу на тестових даних. Це дає змогу *Random Forest* автоматично враховувати внесок кожного дерева в узагальнений прогноз.

4. *Відбір важливих факторів.* *Random Forest* автоматично оцінює вплив кожного фактора на узагальнений прогноз, що далі використовується для відбору впливових факторів.

5. *Виявлення взаємозв'язків.* Під час формування узагальненого прогнозу виявляються взаємозв'язки між факторами, що дає змогу отримати глибше розуміння взаємодії змінних у моделі.

Враховуючи здатність *Random Forest* працювати з великою кількістю факторів, а також виявляти складні взаємозв'язки, цей метод є ефективним інструментом для прогнозування та аналізу даних, а також для відбору важливих факторів та виявлення взаємозв'язків між ними. Для аналізу ефективності методу в контексті розв'язання задачі прогнозування погодинного графіка вироблення електроенергії ФЕС на добу наперед використано блок даних, фрагмент якого подано у табл. 1. Набір даних містить ретроспективні значення впливових факторів за 12 місяців: вимірний обсяг згенерованої електроенергії з погодинним інтервалом «fact_gen» (кВт-год), прогнозовані на добу наперед графіки зміни значень температури повітря «temperature_c» (°C), атмосферного тиску «pressure_in» (inHg), кількості опадів «precip_mm» (мм), вологості «humidity» (в.о.), хмарності «cloud» (в.о) та точки роси «dewpoint_c» (°C). Кожне значення має мітку дати та часу «time».

Таблиця 1

Частина даних з блока даних «ФЕС-1»

time	fact_gen	temperature_c	pressure_in	precip_mm	humidity	cloud	dewpoint_c
2023-06-01 00:00:00	0	10,8	30,08	0	74	3	6,4
2023-06-01 01:00:00	0	10,4	30,08	0	74	0	5,9
2023-06-01 02:00:00	0	9,9	30,08	0	73	1	5,2
2023-06-01 03:00:00	0	9,3	30,08	0	72	0	4,5
2023-06-01 04:00:00	0	8,9	30,09	0	72	0	4,2
2023-06-01 05:00:00	112,4739	9,8	30,09	0	72	0	5
2023-06-01 06:00:00	686,0913	12,7	30,1	0	64	0	6,1
2023-06-01 07:00:00	2334,6	14,7	30,1	0	56	0	6,1
2023-06-01 08:00:00	4260,6	16,5	30,09	0	51	0	6,2
2023-06-01 09:00:00	5816,7	17,8	30,09	0	45	0	5,8
2023-06-01 10:00:00	6857,1	18,8	30,09	0	42	0	5,7
2023-06-01 11:00:00	6994,8	21,6	30,08	0	41	0	7,6
2023-06-01 12:00:00	6990,3	22,5	30,07	0	40	0	8,1
2023-06-01 13:00:00	6987,6	22	30,05	0	39	4	7,6
2023-06-01 14:00:00	6957,9	21,8	30,03	0	40	12	7,5
2023-06-01 15:00:00	6205,5	21,8	30,02	0	40	24	7,7
2023-06-01 16:00:00	4855,5	21,7	30,02	0	41	26	8
2023-06-01 17:00:00	3194,1	21,7	30,02	0	44	14	8,8
2023-06-01 18:00:00	1301,0826	21,1	30,02	0	54	9	11,4
2023-06-01 19:00:00	263,0652	18,2	30,03	0	62	9	10,8
2023-06-01 20:00:00	0	14,6	30,03	0	63	6	7,7
2023-06-01 21:00:00	0	13,8	30,03	0	65	4	7,2
2023-06-01 22:00:00	0	13,2	30,03	0	66	4	7,1
2023-06-01 23:00:00	0	12,8	30,02	0	67	3	6,9

Зазначені метеопараметри для прогнозування генерування ФЕС вибрано на основі статистичних досліджень. Аналіз даних за трирічний період показав, що зміни саме цих даних найбільше корелюють з добовими графіками згенерованої електроенергії. Джерелом прогнозних даних щодо метеопараметрів у точках розміщення ФЕС є погодні сервіси, зокрема *Meteoblue*, *AccuWeather*, *Global Weather API* та ін. Не зважаючи на використання низки сервісів для підвищення достовірності прогнозу метеопараметрів, якість прогнозу є далекою від ідеальної, особливо на території України, де фактично відсутня мережа метеостанцій, оснащених засобами телевимірювання.

Щодо точності вимірюваних значень, зокрема погодинного вироблення електроенергії, температури повітря, швидкості вітру, відносної вологості, то на неї можуть негативно впливати збої в роботі датчиків, первинних перетворювачів та каналів зв'язку, відсутність інтернет-з'єднання,

перерви електропостачання базових станцій мережі 3G/4G, а також інші зовнішні фактори. Тому перед аналізом даних, формуванням навчальної вибірки та навчанням моделей необхідно виконувати двоступеневе оброблення вихідних даних.

На першому етапі необхідно виключити з навчальної вибірки, сформованої на основі датасету «ФЕС-1», заміри обсягів електроенергії та дані про погодні умови, які відповідають аномальним режимам роботи. Для ФЕС до таких режимів належать режими технічних обмежень з боку оператора системи передачі, операторів систем розподілу, диспетчерів електростанції або керувальної організації; режими, пов'язані з аварійними відключеннями основного обладнання; післяаварійні та ремонтні режими тощо.

На другому етапі необхідно позбутись помилкових відхилень (викидів) вимірюваних даних, та даних про погодні умови, отриманих з метеосервісів. Для цього варто проаналізувати результати вимірювання електричних та метеорологічних параметрів й відфільтрувати години, коли система моніторингу працювала некоректно. Такі періоди можна ідентифікувати, використовуючи протоколи функціонування системи моніторингу або характерні ознаки вимірюваних даних. Наприклад, якщо протягом кількох годин поспіль значення параметрів не змінюються або ж відсутні, то це свідчить про збій у системі моніторингу. На цьому ж етапі виміряні метеопараметри зіставляються з отриманими з метеоресурсів (прогнозними на добу наперед). Дані, що мають значні відхилення маркуються, як дані з низькою точністю. Далі, враховуючи природний взаємозв'язок між трендами окремих параметрів, виявляються їхні нетипові зміни. Такі дані також маркуються як дані з низькою точністю.

Після етапу тестування та валідації даних навчальної вибірки, проведено навчання моделі. Дані для навчання та оцінювання його результатів розділено таким чином: 92 % записів (дані за 11 місяців) використано для формування навчальної вибірки, а решту 8 % записів (дані за 1 місяць) використано для тестування моделі. Це дало змогу оцінити адекватність регресійної моделі, сформовану за методом *Random Forest*, на основі метеопараметрів, на яких навчання не виконувалося. Для тестування моделі вибрано проміжок з 1 по 30 червня 2023 року, оскільки в районі розміщення дослідної ФЕС цей період характеризується великою кількістю (більше 60 %) сонячних днів та достатньо стабільними погодними умовами. Разом з тим, в певні дні спостерігається різка зміна хмарності, що є причиною значних та швидких коливань потужності генерування ФЕС (рис. 1) та обсягів вироблення енергії протягом годинних інтервалів. Для об'єктивного оцінювання точності цей період часу виключено з навчальної вибірки та внесено до тестової.

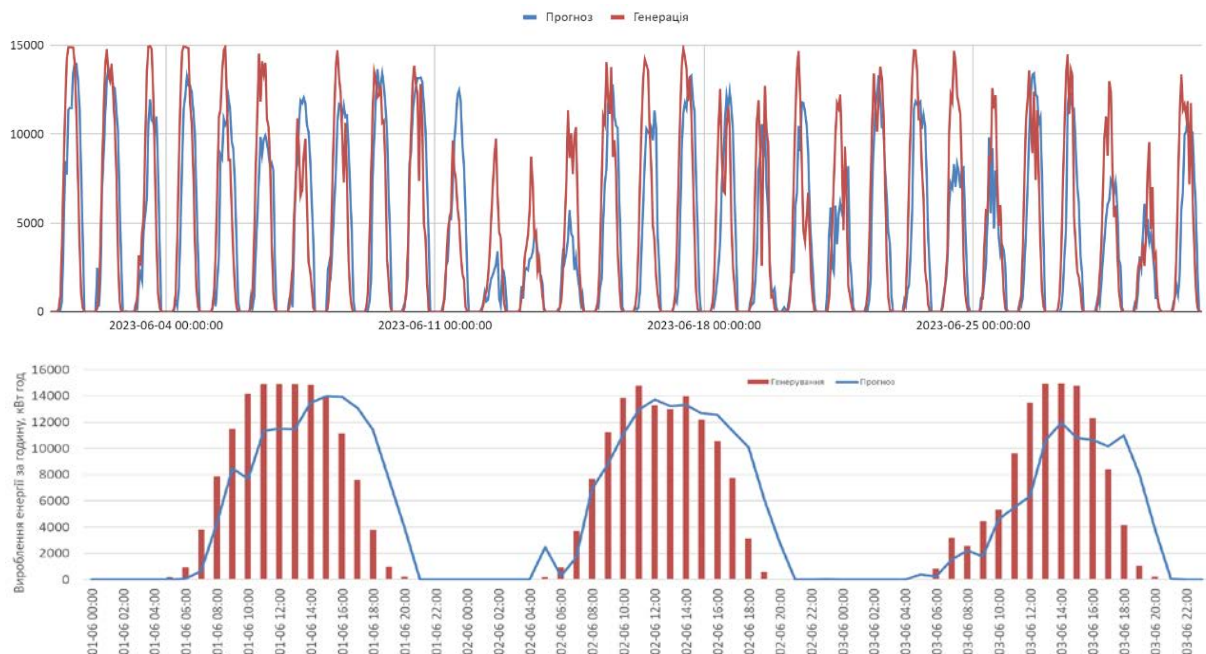


Рис. 1. Результати прогнозування вироблення електроенергії ФЕС.
Обсяг навчальної вибірки обмежено двома місяцями навколо цільового $RSFE_{mid} = 47\%$

Для оцінювання точності прогнозу використано показник Running Sum of Forecast Error (RSFE), який визначається відношенням суми помилок прогнозу за підсумком, що зростає:

$$RSFE = \frac{\sum_i (|W_{Fi} - W_{Oi}|)}{\sum_i (W_{Oi})},$$

де W_{Oi} — вимірне значення відпуску електроенергії за i -ту годину (елемент датасету, оригінал); W_{Fi} — прогнозне значення відпуску електроенергії за i -ту годину. Показник розраховувався для кожної доби тестової вибірки $RSFE_j$ та для вибірки в цілому $RSFE_{mid}$. Аналіз значень $RSFE_j$ дає змогу оцінити максимальну $RSFE_{max}$ та мінімальну $RSFE_{min}$ помилку добового прогнозу вироблення електроенергії для тестової вибірки, а $RSFE_{mid}$ характеризує осереднену помилку прогнозу.

Використовуючи навчальну вибірку, сформовано регресійну модель *Random Forest*, яка дала змогу отримати результати прогнозування графіків генерування, поданих на рис. 1. Для формування моделі використовувалися типові налаштування методу навчання. Алгоритм побудови дерев прийняття рішень ніяк не обмежувався, як і алгоритм визначення та оцінювання впливових факторів для формування регресійних моделей окремих дерев.

Таким чином, для формування прогнозу вироблення електроенергії за кожну годину доби алгоритм навчання використовував всю навчальну вибірку. Міра впливу окремих годинних інтервалів протягом доби на обсяг електроенергії, що виробляється ФЕС протягом конкретної години, визначалася й оцінювалася алгоритмом навчання. За результатами низки розрахунків встановлено, що формування регресійних моделей на основі всього набору даних (рік спостережень) не дає прийнятних результатів (помилки прогнозу $RSFE_{mid} > 50\%$, $RSFE_{min} > 10\%$, $RSFE_{max} > 180\%$). Це пояснюється принциповою відмінністю впливу погодних параметрів, що увійшли до датасету, на генерування ФЕС у різні періоди року. Для отримання помилки, меншої за 50% обсяг датасету для навчання обмежено двома попередніми та двома наступними місяцями відносно періоду прогнозування. Отримані результати не можуть використовуватися на практиці, але дають розуміння суперечливості впливу обсягів навчальної вибірки та необхідності їхньої оптимізації. Після такого обмеження отримано $RSFE_{mid} = 47\%$, $RSFE_{min} = 28\%$, $RSFE_{max} = 86\%$.

Для підтвердження гіпотези про необхідність обмеження блоку даних виконано повторну побудову регресійних моделей на основі вибірки, розширеною до трьох місяців до та після цільового періоду прогнозу. Це забезпечило кращу відповідність між формою прогнозних та вимірних графіків генерування, проте середня помилка очікувано зросла до 51% ($RSFE_{mid} = 51\%$, $RSFE_{min} = 32\%$, $RSFE_{max} = 116\%$). Погіршення точності прогнозу пояснюється відмінністю у характері генерування ФЕС для місяців, що значно віддалені від цільового. Іншими словами, до генерування ФЕС у червні більше подібне генерування ФЕС у квітні та травні, ніж у березні.

Обмеження обсягу навчальної вибірки до одного місяця до та після цільового також призвело до зростання середньої помилки прогнозу ($RSFE_{mid} = 53\%$, $RSFE_{min} = 27\%$, $RSFE_{max} = 132\%$). Таким чином, для блоку даних, що аналізується, оптимальним можна вважати обмеження обсягу навчальної вибірки двома місяцями відносно цільового. Цей висновок, очевидно не є узагальненим, адже залежить від переліку параметрів, включених до блоку даних, їхньої зв'язності та валідності. Однак він вказує один з напрямків поліпшення вихідних даних.

Іншою особливістю генерування ФЕС є їх істотна залежність від часу доби. Адже це визначає кут падіння сонячних променів на поверхню панелі незалежно від інших погодних умов, які впливають на інсоляцію. Тому подальшим етапом оптимізації навчальної вибірки є визначення кількості годин, якою варто обмежити навчання для формування регресійної моделі для прогнозування цільової години доби. Для підтвердження важливості цього фактора перебудовано регресійну модель з попередніми налаштуваннями та додатковим обмеженням блоку даних. Навчання виконувалося за даними глибиною в два місяці, але лише для тої години, для якої формувався прогноз. Наприклад, прогнозуючи вироблення енергії за десяту годину ранку, для навчання модель могла використовувати обмежений датасет за попередні та наступні два місяці, що містив лише дані за десяту годину ранку. Наслідком такого уточнення є значне зменшення помилки прогнозування ($RSFE_{mid} = 22\%$, $RSFE_{min} = 8\%$, $RSFE_{max} = 70\%$, рис. 2).

Враховуючи значний вплив означеного фактора на точність прогнозування вироблення енергії ФЕС виконано низку розрахунків з обмеженням кількості годин, залишених у блоці даних для

формування моделі прогнозування цільової години. Кількість годин варіювалася від нуля до п'яти. Результати оцінювання помилки прогнозу подано у табл. 2.

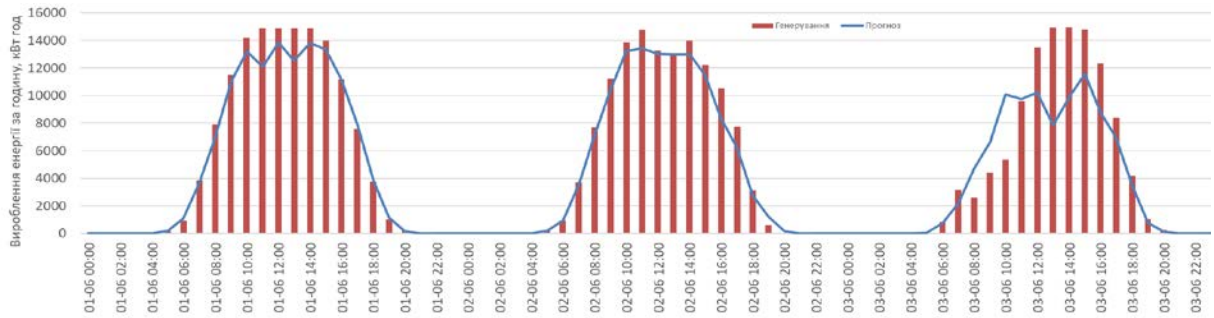


Рис. 2. Результати прогнозування вироблення електроенергії ФЕС. Обсяг навчальної вибірки обмежено цільовою годиною. $RSFE_{mid} = 22\%$

Таблиця 2

Результати оптимізації блоку даних за кількістю годин у навчальній вибірці по відношенню до цільової години прогнозу

Кількість годин доби, які передували цільовій годині і враховані у навчальній вибірці	Кількість годин доби, після цільової години, які враховані у навчальній вибірці	Середня помилка прогнозу генерування за місяць, $RSFE_{mid}$, %	Мінімальна помилка прогнозу генерування за добу, $RSFE_{min}$, %	Максимальна помилка прогнозу генерування за добу, $RSFE_{max}$, %
0	0	22	8	70
1	1	22,1	9	68
2	2	22,4	11	66
3	3	23,1	12	65
4	4	24,0	14	63
5	5	25,4	16	62

З результатів розрахунку видно, що збільшення кількості годин, що передують цільовій годині у навчальній вибірці збільшує середню помилку прогнозу за місяць, проте зменшує її максимальну оцінку, вирівнюючи ефективність регресійних залежностей, отриманих *Random Forest* для різних погодних умов. Подальше збільшення кількості годин у навчальній вибірці, відносно цільової години, фактично не зменшує значення $RSFE_{max}$, підвищуючи мінімальну й середню оцінки помилки прогнозу. Таким чином, розширення вибірки виявляється недоцільним.

Іншим шляхом для врахування особливостей роботи ФЕС у години мінімального та максимального генерування, є застосування фільтрів за обсягом погодинного вироблення електроенергії. Так, в години мінімального генерування істотно зростає невизначеність впливу характеристик погодних умов, через значний вплив факторів, які складно врахувати. Зокрема, відмінність у параметрах сонячних панелей та інверторного обладнання, різні кути нахилу панелей та різна їхня орієнтація в межах одної ФЕС тощо. Таким чином, навіть у разі підтвердженої валідності параметрів, які внесені до блоку даних для ранішніх та вечірніх годин, їхні залежності можуть істотно суперечити співвідношенням, отриманим у години середньої та максимальної інсоляції.

З іншого боку, значні «викиди» генерування, зафіксовані та підтвержені в години максимальної інсоляції, на практиці можуть бути наслідком впливу одного, або кількох параметрів, які не вимірюються в межах ФЕС, або валідність яких не може бути підтверджена. Таким чином, використання в блоці даних і записів, що відповідають граничним значенням погодинного генерування в межах доби, може сприяти погіршенню якості прогнозування.

Для підтвердження означених положень виконано низку розрахунків з формування регресійних моделей на основі датасетів, записи у яких відфільтровано за ознакою співвідношення зафіксованого значення погодинного генерування ФЕС з його середнім значенням за добу. Таким чином до блоку даних, що застосовувався для формування регресійної моделі прогнозування обсягу генерування ФЕС для певної години включалися дані за два місяці до та після цільового місяця, з урахуванням всіх погодинних записів, крім тих, що відфільтровані за ознакою граничних мінімального та максимального значень генерування. Результати розрахунків подано у табл. 3.

Таблиця 3

Результати оптимізації блоку даних за максимальним та мінімальним погодинним генеруванням ФЕС, яке враховувалося у добовому графіку

Відношення мінімального генерування ФЕС до осередненого за добу, %	Відношення максимального генерування ФЕС до осередненого за добу, %	Середня помилка прогнозу генерування за місяць, $RSFE_{mid}$, %	Мінімальна помилка прогнозу генерування за добу, $RSFE_{min}$, %	Максимальна помилка прогнозу генерування за добу, $RSFE_{max}$, %
0	0	47	28	86
10	190	42	30	64
20	180	37	29	58
30	170	31	24	48
40	160	27	20	41
50	150	22	17	34

Таким чином, застосування попередньої фільтрації даних датасету за критерієм мінімального та максимального генерування дав змогу отримати результати, що за осередненою оцінкою помилки прогнозу $RSFE_{mid}$ відповідають попереднім. Проте, отримані регресійні моделі є значно стійкішими, що проявляється у меншому розкиді оцінок помилки прогнозування (див. табл. 3): оцінка $RSFE_{min}$ зросла, порівняно з найкращим випадком з табл. 2, але оцінка $RSFE_{max}$ істотно зменшилася для тих самих вихідних даних для прогнозування. Графічне представлення результатів прогнозування подано на рис. 3.

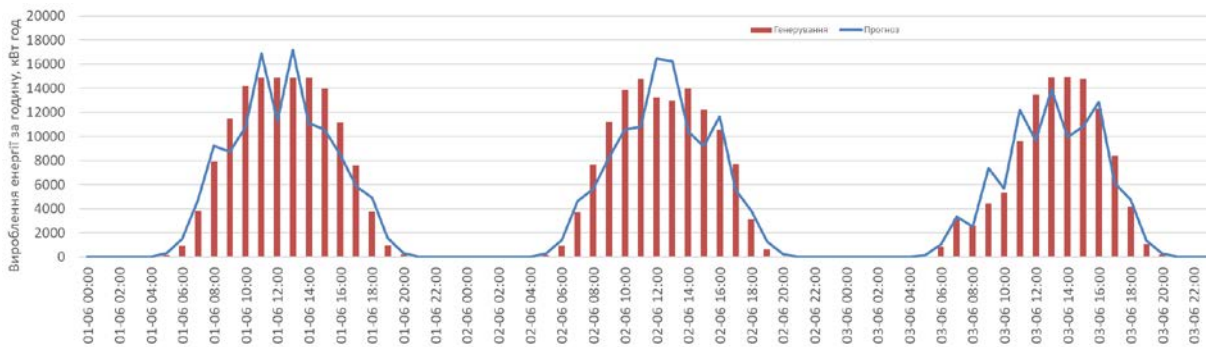


Рис. 3. Результати прогнозування вироблення електроенергії ФЕС.

Обсяг навчальної вибірки обмежено фільтрацією за обсягом генерування. $RSFE_{mid} = 22\%$

Аналізуючи результати прогнозування генерування (рис. 3) виявлено, що вони мають неідеальну куполоподібну форму, яка є характерною для вибраного періоду прогнозування. Це може вказувати на наявність некерованих аномалій в даних, що впливають на форму прогнозного графіка. Для згладжування таких непрямолинійних змін можна застосувати фільтр Савицького-Голея [16]. Цей фільтр використовується для згладжування кривих шляхом видалення шуму та непродуктивних варіацій, які можуть спотворювати результати аналізу. В результаті застосування цього фільтра для кінцевої обробки результатів прогнозування, отримано певне поліпшення результатів прогнозування, що проявляється у зменшенні помилки прогнозування: $RSFE_{mid} = 21\%$.

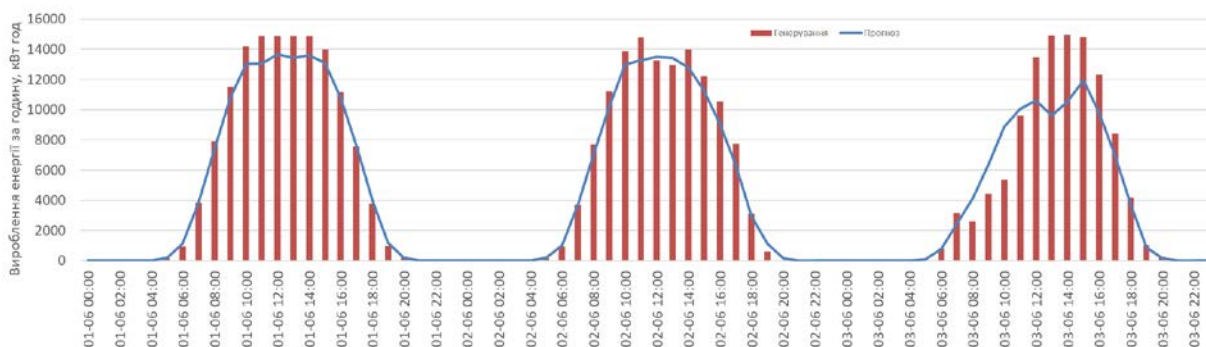


Рис. 4. Результати прогнозування вироблення електроенергії ФЕС.

До результатів прогнозування застосовано фільтр Савицького-Голея. $RSFE_{mid} = 21\%$

Висновки

Отже, застосування методу *Random Forest* виявило певний потенціал для прогнозування. Його здатність ідентифікувати складні залежності, нелінійні співвідношення та сезонні патерни, враховуючи зовнішні фактори забезпечує його корисність в широкому спектрі сценаріїв прогнозування. Проте аналіз результатів вказує на значні помилки в прогнозуванні протягом тестового періоду. Це свідчить про необхідність попередньої оптимізації блоку даних для ефективного використання методу. Зокрема, важливо враховувати сезонні зміни та інші метеорологічні фактори, що може суттєво вплинути на точність прогнозу.

Варто зазначити, що ефективність остаточного прогнозу тісно пов'язана як з якістю вихідних даних для навчання моделі, так і з якістю прогнозу даних метеопараметрів. Для подальшого поліпшення результатів, варто розглянути застосування додаткових метеопараметрів, які надають спеціальні метеосервіси, зокрема, DNI (direct normal irradiance), DHI (diffuse horizontal irradiance), GHI (global horizontal irradiance), а також залучити методи валідації метеопараметрів на основі даних від різних постачальників. Більше того, доцільно детальніше дослідити питання оптимального співвідношення глибини навчальної вибірки та налаштувань фільтрів за обсягом погодинного генерування для врахування сезонності та підвищення точності прогнозу.

Для підвищення точності алгоритмів прогнозування перспективним видається застосування напрацьованих Мішеля Талагранна (зокрема, методу «ланцюгування») [17]. Його ключовою перевагою є здатність не лише надавати оцінки очікуваного, але й окреслювати верхні та нижні межі для цих величин, що може стати цінним інструментом для оцінювання ризиків, пов'язаних з прогнозуванням. А це є актуальним фактором в енергетичному секторі.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] A. Loureno, et al. "Comparison of forecasting models for photovoltaic power generation," *Energy Conversion and Management*, no. 118, pp. 404-418, 2016.
- [2] Peter J. Brockwell and Richard A. Davis, *Time Series: Theory and Methods*, 2016. [Electronic resource]. Available: <https://pdfarchived.net/list/time-series-theory-and-methods-peter-j-brockwell-4913414>.
- [3] Robert H. Shumway and David S. Stoffer, *Time Series Analysis and Its Applications: With R Examples*, 2017. [Electronic resource]. Available: <http://pzs.dstu.dp.ua/DataMining/times/bibl/TimeSeries.pdf>.
- [4] G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, no. 50, pp. 159-175. 2003. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0).
- [5] S. Borchani, et al. "Short-term solar power forecasting using machine learning techniques," *Renewable Energy*, no. 116, Part A, pp. 729-743, 2018.
- [6] A. C. Cadena, et al. "Weather Forecasting for Photovoltaic Power Prediction Using Machine Learning Techniques," *Energies*, no. 11(6), pp. 1362, 2018.
- [7] R. J. Hyndman, and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts. 2018. [Electronic resource]. Available: <https://www.scirp.org/reference/referencespapers?referenceid=2849375>.
- [8] C. Zhang, Y. Guo, M. Li "A review of the development and application of artificial neural network models [J]," *Computer Engineering and Applications*, no. 57(11), pp. 57-69, 2021.
- [9] Yi Zhou, et al. "Prediction of photovoltaic power output based on similar day analysis, genetic algorithm and extreme learning machine," *Energy*, no. 204, pp. 78-94, 2020.
- [10] D. Infield, and M. O'Malley, "A probabilistic forecast methodology for the management of renewable energy: The application of small-scale solar power," *IEEE Transactions on Power Systems*, no. 22(3): 1147, pp. 1156, 2007.
- [11] Mittal, Amit Kumar, Kirti Mathur, and Shivangi Mittal, "A review on forecasting the photovoltaic power using machine learning," *Journal of Physics: Conference Series*, vol. 2286, no. 1. IOP Publishing, 2022.
- [12] Ibrahim I. Anwar, M. J. Hossain, and Benjamin C. Duck, "An optimized offline random forests-based model for ultra-short-term prediction of PV characteristics," *IEEE Transactions on Industrial Informatics*, no. 16.1, pp. 202-214, 2019.
- [13] Andreas C. Müller and Sarah Guido, *Introduction to Machine Learning with Python: A Guide for Data Scientists*, 2016.
- [14] Kumar Abhishek, et al. "A review and analysis of forecasting of photovoltaic power generation using machine learning," *International Conference on Management Science and Engineering Management*. Cham: Springer International Publishing, 2022.
- [15] Lei Wen, and Xiaoyu Yuan. "Forecasting CO₂ emissions in Chinas commercial department, through BP neural network based on random forest and PSO," *Science of the Total Environment*, no. 718, pp. 137-194, 2020. ISSN 0048-9697. <https://doi.org/10.1016/j.scitotenv.2020.137>.
- [16] A. Savitzky, M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Analytical Chemistry*, no.36(8), pp. 1627-1639, 1964.
- [17] Michel Talagrand, *The Generic Chaining: Upper and Lower Bounds of Stochastic Processes*, Springer-Verlag, 2005, 222 pp.

Рекомендована кафедрою електричних станцій та систем ВНТУ

Стаття надійшла до редакції 23.10.2024

Кулик Володимир Володимирович — д-р техн. наук, доцент, професор кафедри електричних станцій та систем, e-mail: volodymyrvkulyk@gmail.com ;

Затхей Максим Вікторович — аспірант кафедри електричних станцій та систем, e-mail: maxzathey@gmail.com .

Вінницький національний технічний університет, Вінниця

V. V. Kulyk¹
M. V. Zatkhei¹

Improving the Accuracy of the Forecast of Electricity Production by Photovoltaic Power Station Based on the Random Forest Method

¹Vinnitsia National Technical University

The work examines the methods used to forecast energy production by Photovoltaic Power Station (PPS), as well as ways to improve the accuracy of the forecast to optimize the structure of the electricity balance in the power system. The research is aimed at identifying effective forecasting approaches and algorithms, assessing their accuracy and reliability. Based on the results of the study, a combination of regression analysis and machine learning methods is proposed, which provides acceptable forecast accuracy for planning power reserves in the power system. The Random Forest method was used as a basis, as it ensures adaptability to the specifics of energy generation by PPS in different regions of Ukraine and in different periods of the year. To increase the efficiency of machine learning, an algorithm for pre-filtering datasets using autoregression and moving average methods was proposed. This allows better preparation of input data for further use in forecasting, ensuring smoothing of time series and elimination of random fluctuations that can negatively affect forecast accuracy. The use of pre-filtering methods allows to identify the main regularities in the data, which, in turn, increases the accuracy of machine learning models. The Random Forest method was not chosen by chance: it is well suited for forecasting tasks where many different factors need to be taken into account, which may affect the results differently depending on time and region. This is especially important in the case of forecasting the energy generation of solar power plants, where production is significantly affected by variables such as cloud cover, temperature, seasonality, etc. Using Random Forest allows you to take into account non-linear dependencies and interactions between factors, which helps to increase the accuracy of forecasts. The use of Random Forest as the main machine learning algorithm is due to its flexibility and ability to adapt to different conditions, which allows you to effectively take into account the features of energy generation in different regions and during different seasons.

Keywords: photovoltaic power station, generation, forecasting, regression analysis, machine learning, data analysis.

Kulyk Volodymyr V. — Dr. Sc. (Eng.), Associate Profesor, Professor of the Chair of Power Plants and Systems, e-mail: volodymyrvkulyk@gmail.com ;

Zathey Maksym V. — Post-Graduate Student of the Chair of Power Plants and Systems, e-mail: maxzathey@gmail.com