УДК 621.311

Є. А. Тептя
В. О. Комар

ЗАСТОСУВАННЯ БАЄСІВСЬКОЇ ОЦІНКИ РИЗИКУ ДЛЯ ВИЗНАЧЕННЯ АНОМАЛІЙ В ФАКТИЧНИХ ПОГОДИННИХ ГРАФІКАХ ГЕНЕРУВАННЯ ФЕС

Вінницький національний технічний університет;

Анотація

У роботі запропоновано підхід до виявлення аномалій у графіках генерування фотоелектричних станцій (ФЕС) на основі баєсівської-оцінки ризику недовиробітку методом Монте-Карло для марковських процесів (МСМС). Підхід поєднує фізичну модель очікуваної потужності, побудовану на сонячній геометрії та градієнтному бустингу, з ймовірнісним аналізом залишків через розподіл Стьюдента. Алгоритм враховує тільки випадки недовиробітку, оцінює ризик аномалії як ймовірність потрапляння у хвіст розподілу і додатково агрегує його у часі, щоб відрізнити короткі хмарні затьмарення від справжніх тривалих відхилень.

Ключові слова: сонячна електростанція, аномалія, баєсівський аналіз, МСМС, аналіз залишків, диспетчерські обмеження, очищення даних, прогнозування генерації

Abstract

This paper proposes an approach to detecting anomalies in the power generation curves of photovoltaic power plants (PVPP) based on a Bayesian assessment of the risk of under-generation using the Monte Carlo method for Markov processes. The approach combines a physical model of expected power output, based on solar geometry and gradient boosting, with a probabilistic analysis of residuals using the Student's t-distribution. The algorithm considers only cases of underproduction, estimates the risk of an anomaly as the probability of falling into the tail of the distribution, and further aggregates it over time to distinguish short-term cloud cover events from genuine long-term deviations.

Keywords: solar power plant, anomaly, Bayesian analysis, MCMC, residual analysis, dispatch constraints, data cleaning, generation forecasting

Вступ

Точне прогнозування виробітку сонячних електростанцій (ФЕС) важливе для надійної роботи енергосистеми з великою часткою відновлюваних джерел. Проте моделі прогнозу погано працюють, коли в історичних даних є штучні провали потужності, спричинені або примусовим обмеженням з боку оператора мережі, або технічними проблемами на станції. Такі провали зміщують прогноз і знижують його якість. Тому виникає практична задача – очистити історичні графіки виробітку та оцінити, наскільки достовірною є кожна година спостережень. Можна виділити п'ять основних причин аномальної поведінки графіка. Перша – диспетчерські обмеження потужності, коли оператор мережі змушує станцію працювати нижче можливого. Друга – несправності обладнання, а саме деградація чи відмова інверторів, окремих рядків модулів, систем комутації тощо. Третя – обмеження по верхній межі, коли інвертор не може видати всю наявну потужність постійного струму. Четверта – втрати зв'язку, через які з'являються пропуски та артефакти в телеметрії. П'ята – невизначеність погоди, коли прогноз сонячної радіації не збігається з фактичними хмарами. Оскільки кожна причина аномалії має свій фізичний механізм, то коректна діагностика вимагає не просто фіксації відхилення від прогнозу, а й класифікації його причини.

Результати дослідження

В роботі пропонується алгоритм, який виконує сім послідовних кроків. Спочатку п'ятихвилинні виміри потужності зводяться до погодинних, щоб синхронізуватись із кадансом погодинних метеоданих. Далі для кожної години розраховуються параметри сонячної геометрії – висота сонця над горизонтом, азимут, опромінення в площині панелі. Це виконується засобами бібліотеки *Pvlib* [1]. На цій основі будується модель очікуваної потужності *P_expected* як функція погоди, геометрії та часу. Для апроксимації використовується градієнтний бустинг *LightGBM* [2]. Далі обчислюється залишок *r_t = P_actual – P_expected*, тобто різниця між реальною та очікуваною потужністю за конкретну годину (рис. 1). На наступному кроці залишки аналізуються байєсівськими методами через МСМС [3], після чого ризик аномалії агрегується в часі та інтервал отримує одну з міток класифікації.

У роботі використано метеорологічні параметри – *GHI*, *DNI*, *DHI*, температуру повітря та швидкість вітру – із двох незалежних джерел, *Meteoblue* та *Weatherbit*, а як фактична генерація - п`ятихвилинні графіки потужності трикіловатної ФЕС Вінницького національного технічного університету (ВНТУ) за період з грудня 2019 року по вересень 2025 року, що становить близько 288 тис. записів. Інтервали з провалами зв'язку виявляються окремо і не беруть участі у навчанні моделі очікуваної потужності.

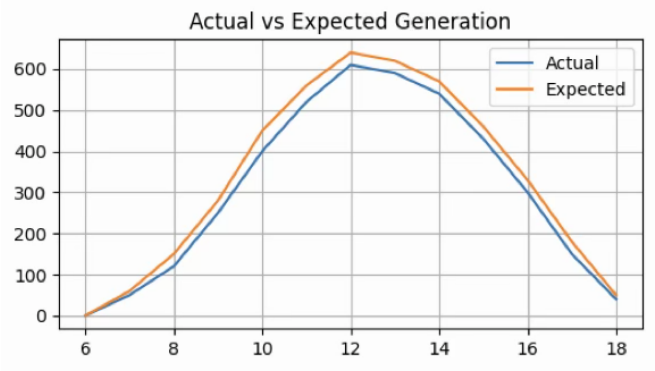

Рис. 1. Аналіз залишків між фактичною та очікуваною потужністю станції

Залишок r_t є базою для виявлення аномалій. Замість фіксованого порогу типу $|r_t| > k \cdot \sigma$ застосовано байєсівський підхід [4], який враховує невизначеність самої оцінки параметрів, дає ймовірнісну характеристику ризику, стійкий до викидів і коректно поводить себе на коротких вибірках. Розподіл залишків моделюється Стьюдент-Т розподілом $r_t \sim StudentT(\mu, \sigma, \nu)$, його «важкі хвости» добре описують рідкісні різкі провали потужності, типові для реальних ФЕС.

Параметри $\theta = (\mu, \sigma, \nu)$ оцінюються методом MCMC за теоремою Байєса $P(\theta \mid дані) \propto P(дані \mid \theta)\, P(\theta)$, що дає не одне значення, а цілий апостеріорний розподіл. Важлива деталь: враховуються лише від'ємні залишки $neg_res_t = max(0, -r_t)$. Це фізично обґрунтовано тим, що ФЕС принципово не може видати потужності більше, ніж дозволяє сонячна радіація – то ж додатні залишки трактуються як шум вимірювань, а аномаліями вважаються тільки випадки недовиробітку. Ризик кожної години оцінюється як ймовірність опинитись у хвості розподілу: $Risk(t) = 1 - F_StudentT(neg_res_t; \hat{\mu}, \hat{\sigma}, \hat{\nu})$. Чим вищий ризик – тим імовірніше, що ця година є аномалією. Одна година з високим ризиком сама по собі ще нічого не доводить – часто це звичайна хмара, що пройшла над станцією. Натомість, якщо ризик залишається високим декілька годин поспіль, це вже привід звернути увагу на дану аномалію. Тому миттєвий ризик додатково агрегується в часі: оцінюється не тільки інтенсивність недовиробітку, а також його тривалість. Часова агрегація ризику для відокремлення короткочасних збурень від тривалих аномалій наведена на рис. 2.

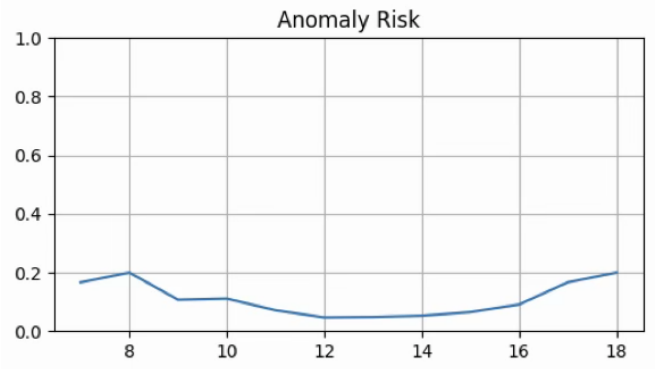

Рис. 2. Часова агрегація ризику для відокремлення короткочасних збурень від тривалих аномалій

На основі агрегованого ризику та супровідних погодних умов кожна година розподіляється до одного з чотирьох класів: *NORMAL* – звичайна робота станції; *CURTAILMENT* – диспетчерське обмеження за хороших погодних умов; *POSSIBLE_FAULT* – підозра на несправність обладнання;

WEATHER_UNCERTAIN – недовиробіток, який зумовлений невизначеністю прогнозу хмарності [5, 6].

Наукова новизна роботи полягає в тому, що тут вперше поєднано фізичну модель сонячної електростанції з байєсівською оцінкою ризику в одному алгоритмі. Виконано асиметричний аналіз лише від'ємних залишків, що відповідає фізиці фотоелектричної генерації. Ризик агрегується в часі ймовірнісно, а не як просте середнє значення. Для опису залишків ФЕС вперше застосовано *Student-T* розподіл у поєднанні з *MCMC*-семплюванням.

Практична цінність підходу полягає в можливості очистити дані перед навчанням моделей прогнозування, зменшити їх систематичне зміщення, виявити епізоди диспетчерського обмеження та підвищити якість прогнозу генерації.

Перевагами такого підходу є зрозумілість результату (для кожної години видно, чому її помічено як аномалію), ймовірнісна оцінка впевненості, фізична обґрунтованість і стійкість до шуму.

В той же час метод має певні обмеження: точність залежить від якості прогнозу погоди; відсутні еталонні дані для абсолютної перевірки результату; історичний період даних обмежений; повністю довести епізод диспетчерського обмеження без журналів оператора неможливо.

Висновки

Підсумовуючи вищезазначене, розроблений підхід поєднує фізичну модель ФЕС із байєсівським аналізом залишків і дозволяє ефективно очищати історичні графіки генерації та формувати якісні вибірки для навчання моделей прогнозування. Метод апробовано на реальних даних ФЕС ВНТУ і може бути перенесений на станції більшої потужності за умови наявності якісних метеоданих. Даний підхід можна поширювати на ієрархічні моделі для групи ФЕС у межах однієї мережі та інтегрувати з прихованими марковськими моделями для більш тонкої класифікації режимів їх роботи.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Holmgren W. F., Hansen C. W., Mikofski M. A. pvlib python: a python package for modeling solar energy systems. Journal of Open Source Software. 2018. Vol. 3, No. 29. P. 884. URL: https://doi.org/10.21105/joss.00884 (дата звернення: 07.05.2026).

2. Ke G., Meng Q., Finley T. et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems. 2017. Vol. 30. P. 3146-3154. URL: https://proceedings.neurips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html (дата звернення: 01.05.2026).

3. Abril-Pla O., Andreani V., Carroll C. et al. PyMC: a modern, and comprehensive probabilistic programming framework in Python. PeerJ Computer Science. 2023. Vol. 9. e1516. URL: https://doi.org/10.7717/peerj-cs.1516 (дата звернення: 11.05.2026).

4. Gelman A., Carlin J. B., Stern H. S. et al. Bayesian Data Analysis. 3rd ed. Boca Raton: CRC Press, 2013. 675 p. URL: https://doi.org/10.1201/b16018 (дата звернення: 12.05.2026).

5. Mellit A., Tina G. M., Kalogirou S. A. Fault detection and diagnosis methods for photovoltaic systems: A review. Renewable and Sustainable Energy Reviews. 2018. Vol. 91. P. 1-17. URL: https://doi.org/10.1016/j.rser.2018.03.062 (дата звернення: 11.05.2026).

6. Bright J. M., Killinger S., Lingfors D., Engerer N. A. Improved satellite-derived PV power nowcasting using real-time power data from reference PV systems. Solar Energy. 2018. Vol. 168. P. 118-139. URL: https://doi.org/10.1016/j.solener.2017.10.091 (дата звернення: 27.04.2026).

7. Hoffman M. D., Gelman A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research. 2014. Vol. 15. P. 1593-1623. URL: https://jmlr.org/papers/v15/hoffman14a.html (дата звернення: 03.05.2026).

Тептя Євгеній Андрійович – аспірант кафедри електричних станцій та систем, факультет електроенергетики та електромеханіки, Вінницький національний технічний університет, м. Вінниця, e-mail: genyashkat@gmail.com;

Комар Вячеслав Олександрович – доктор техн. наук, професор, завідувач кафедри електричних станцій та систем, факультет електроенергетики та електромеханіки, Вінницький національний технічний університет, м. Вінниця, e-mail: kvo1976@ukr.net

Yevhenii A. Teptia – Postgraduate student of the Department of Electrical Power Plants and Systems, Faculty of Electrical Power Engineering and Electromechanics, Vinnytsia National Technical University, Vinnytsia, e-mail:genyashkat@gmail.com;

Viacheslav O. Komar – Doctor of Technical Sciences, Professor, Head of the Department of Power Plants and Systems, Faculty of Electrical Power Engineering and Electromechanics, Vinnytsia National Technical University, Vinnytsia, e-mail: kvo1976@ukr.net