

ИНФОРМАТИКА COMPUTER SCIENCE

IRSTI 28.23.37; 44.41.35

<https://doi.org/10.51889/2959-5894.2026.95.3.012>

A. Abdildayeva¹, G. Taganova^{2*}, A. Tazhibeyev³, A. Kural³, S. Askarov³

¹International Engineering and Technology University, Almaty, Kazakhstan

²Astana International University, Astana, Kazakhstan

³UMAY R&D LLP, Astana, Kazakhstan

*e-mail: guldana.kileuzhanova@gmail.com

LEAKAGE-CONTROLLED RESIDUAL-CALIBRATED HYBRID MODEL FOR ONE-HOUR-AHEAD PHOTOVOLTAIC GENERATION FORECASTING

Abstract

This experimental research study evaluates a leakage-controlled residual-calibrated hybrid model for sequential one-hour-ahead photovoltaic generation forecasting during active station-hours. A source audit of operational records from three solar power plants identified 28,296 raw hourly rows but only 11,736 unique station-hour keys; 16,560 redundant source rows were collapsed only when known plan and actual values were identical, and no conflicting known plan/actual values were found. After restricting supervised modelling to unique station-hours with positive planned generation and known actual output, 4,848 observations remained and were chronologically divided into 3,038 training, 819 calibration, and 991 held-out test observations. The proposed Residual-Calibrated Neural Hybrid (RCNH) combines a two-hidden-layer Tabular MLP with a HistGradientBoosting residual calibrator trained exclusively on the calibration period. On the held-out test period, RCNH achieved MAE = 2,591.80 kWh, RMSE = 3,906.53 kWh, and $R^2 = 0.908609$; 95% three-day block-bootstrap intervals were 2,372.75-2,961.81 kWh, 3,537.16-4,505.29 kWh, and 0.848489-0.932518, respectively. Against the train-only Tabular MLP, RCNH reduced MAE by 11.36% (95% block-bootstrap interval 6.98-14.95%; Holm-adjusted two-sided $p = 0.003998$). However, a matched-label-budget diagnostic favored an MLP refitted on train plus calibration data (MAE = 2,464.63 kWh), and rolling-origin validation showed substantially stronger temporal stability for tree ensembles, with ExtraTrees attaining mean RMSE = 3,064.51 kWh versus 6,106.74 kWh for RCNH. These results show that residual calibration is competitive and beneficial in the final held-out regime but is not uniformly dominant; its effectiveness depends on calibration-window representativeness and temporal regime. The contribution is therefore an experimentally validated, leakage-controlled forecasting protocol and a quantified assessment of where residual calibration helps and where its robustness is limited.

Keywords: photovoltaic generation; one-hour-ahead forecasting; data leakage; residual calibration; chronological validation; block bootstrap; machine learning; experimental evaluation

А.А. Абдилдаева¹, Г. Таганова², А.А. Тажибаев³, А.Ж. Құрал³, С.С. Аскаров³
¹Халықаралық инженерлік-технологиялық университеті, Алматы қ., Қазақстан
²Астана халықаралық университеті, Астана қ., Қазақстан
³UMAY R&D ЖШС, Астана қ., Қазақстан

ФОТОЭЛЕКТРЛІК ГЕНЕРАЦИЯНЫ БІР САҒАТ БҰРЫН БОЛЖАУҒА АРНАЛҒАН ДЕРЕКТЕРДІҢ АҒЫП КЕТУІН БАҚЫЛАЙТЫН ЖӘНЕ ҚАЛДЫҚТАРДЫ КАЛИБРЛЕЙТІН ГИБРИДТІ МОДЕЛЬ

Аңдатпа

Бұл ғылыми-эксперименттік зерттеуде белсенді сағаттардағы фотоэлектрлік генерацияны бір сағат алдын ала болжауға арналған деректердің ағып кетуін бақылайтын қалдықты-калибрленген гибриді модель бағаланды. Үш күн электр станциясының 28 296 бастапқы сағаттық жазбасы аудиттен өткізіліп, 11 736 бірегей «станция-сағат» кілті анықталды; бірдей мәндері бар 16 560 қайталама жазба біріктірілді. Жоспарлы генерациясы оң және нақты генерациясы белгілі 4 848 бірегей белсенді бақылау 3 038 оқыту, 819 калибрлеу және 991 тәуелсіз тест бақылауына хронологиялық түрде бөлінді. RCNN моделі екі жасырын қабатты Tabular MLP мен калибрлеу кезеңінің қалдықтарында оқытылған HistGradientBoosting калибраторын біріктіреді. Тест кезеңінде RCNN MAE = 2 591,80 кВт·сағ, RMSE = 3 906,53 кВт·сағ және $R^2 = 0,908609$ нәтижелерін көрсетті. Оқыту кезеңімен ғана үйретілген MLP-мен салыстырғанда MAE 11,36%-ға төмендеді; үш күндік block-bootstrap үшін 95% аралығы 6,98-14,95%, Holm түзетуінен кейінгі екіжақты $p = 0,003998$. Алайда train+calibration деректерімен қайта оқытылған MLP MAE = 2 464,63 кВт·сағ нәтижесін көрсетті, ал rolling-origin тексеруде ағаштық ансамбльдер анағұрлым тұрақты болды. Сондықтан RCNN бәсекеге қабілетті және белгілі бір тест режимінде тиімді, бірақ оның артықшылығы калибрлеу терезесінің репрезентативтілігі мен уақыттық режимге тәуелді.

Түйін сөздер: фотоэлектрлік генерация; бір сағат алдын ала болжау; деректердің ағып кетуі; қалдықтарды калибрлеу; хронологиялық валидация; block bootstrap; машиналық оқыту; эксперименттік бағалау

А.А. Абдилдаева¹, Г. Таганова², А.А. Тажибаев³, А.Ж. Құрал³, С.С. Аскаров³
¹Международный инженерно-технологический университет, г.Алматы, Казахстан
²Международный университет Астана, г.Астана, Казахстан
³ТОО Umay R&D, г.Астана, Казахстан

ГИБРИДНАЯ МОДЕЛЬ С КОНТРОЛЕМ УТЕЧКИ ДАННЫХ И КАЛИБРОВКОЙ ОСТАТКОВ ДЛЯ ПРОГНОЗИРОВАНИЯ ФОТОЭЛЕКТРИЧЕСКОЙ ГЕНЕРАЦИИ НА ОДИН ЧАС ВПЕРЕД

Аннотация

В данной научно-экспериментальной работе выполнена оценка гибридной модели с контролем утечки данных и калибровкой остатков для последовательного прогнозирования фотоэлектрической генерации на один час вперед в активные часы. Аудит эксплуатационных данных трех солнечных электростанций показал, что 28 296 исходных почасовых строк соответствуют 11 736 уникальным наблюдениям «станция-час»; 16 560 повторных строк с идентичными известными значениями были консолидированы. После ограничения области моделирования наблюдениями с положительной плановой и известной фактической генерацией получено 4 848 уникальных активных наблюдений, хронологически разделенных на 3 038 обучающих, 819 калибровочных и 991 независимое тестовое наблюдение. Модель RCNN объединяет двухслойный Tabular MLP и калибратор HistGradientBoosting, обучаемый только на остатках калибровочного периода. На тестовом периоде получены MAE = 2 591,80 кВт·ч, RMSE = 3 906,53 кВт·ч и $R^2 = 0,908609$. Относительно MLP, обученного только на train-периоде, MAE снижена на 11,36%; 95%-й интервал трехдневного блочного bootstrap составляет 6,98-14,95%, скорректированное по Holm двустороннее $p = 0,003998$. Вместе с тем диагностический MLP, переобученный на train+calibration, достиг MAE = 2 464,63 кВт·ч, а rolling-origin валидация показала более высокую временную устойчивость древесных ансамблей. Следовательно, RCNN является конкурентоспособной и эффективной в конкретном отложенном режиме, но не демонстрирует универсального превосходства; ее результат зависит от репрезентативности калибровочного окна и временного режима.

Ключевые слова: фотоэлектрическая генерация; прогнозирование на один час вперед; утечка данных; калибровка остатков; хронологическая валидация; блочный bootstrap; машинное обучение; экспериментальная оценка

Introduction

Main provisions. The experimental unit is the unique station-hour, with duplicate source exports consolidated before feature engineering. The study evaluates leakage-controlled sequential one-hour-ahead photovoltaic forecasting during active hours using only information available at the forecast origin. On the held-out period, the Residual-Calibrated Neural Hybrid (RCNH) improved the train-only MLP, but matched-label-budget and rolling-origin analyses showed that this advantage was not universal. Overall, model ranking depended on calibration-window representativeness, information budget, and temporal regime. Operational photovoltaic forecasting supports dispatch, deviation monitoring, reserve planning, and integration of variable renewable generation. The task is complicated by non-stationarity across hours, seasons, plant scales, and operating regimes [1-7]. Forecasting modules can also contribute to broader digital-twin and monitoring architectures for energy systems and photovoltaic installations [8,9]. In industrial data, repeated exports and variables whose availability at prediction time is uncertain can introduce data leakage and inflate apparent predictive accuracy [10]. This study addresses two risks that can inflate apparent forecasting accuracy: data leakage from current or future information and the treatment of repeated source rows as independent observations. Accordingly, modelling is restricted to unique active station-hours and historical predictors aligned to exact earlier timestamps.

The research question is whether residual calibration can improve a leakage-controlled neural forecast on a genuinely later period and whether any gain remains credible under time-aware uncertainty analysis and temporal robustness checks.

The objective is to develop and experimentally evaluate RCNH for one-hour-ahead photovoltaic generation forecasting. The study audits the operational data layer, constructs prediction-time admissible features, compares RCNH with operational, linear, tree-based, boosting, and neural baselines, and evaluates uncertainty, matched information budgets, rolling-origin robustness, and ablation effects. The working hypothesis is that residual calibration can improve a primary MLP when the calibration period is representative of the subsequent regime, while strong tabular models may be more stable under temporal shift. Prior photovoltaic forecasting studies cover physical, statistical, machine-learning, and hybrid approaches [3-7]. For temporally ordered forecasting problems, validation must respect chronology and dependence rather than rely uncritically on random resampling [11,12]. For medium-sized tabular datasets, linear models, Random Forest, ExtraTrees, gradient boosting, and XGBoost remain strong benchmarks [13-17]. The gap addressed here is methodological: duplicate source rows, undefined information availability, calibration/test contamination, and reliance on a single split can lead to optimistic conclusions. The proposed protocol therefore combines station-hour auditing, strict feature availability, a dedicated calibration period, an untouched test period, and temporal robustness diagnostics.

Research methodology

Data audit and experimental domain

The operational archive contains hourly plan-fact generation and several daily or event-based tables. Only variables with defensible availability at the forecast origin were retained. Same-day finalized KPIs, curtailment aggregates, maintenance and defect records were excluded when an hourly as-of join or availability timestamp could not be verified. A source audit reduced 28,296 raw hourly rows to 11,736 unique station-hour keys by consolidating only identical repeated values; no conflicting known plan/actual values were found. Among 9,792 unique station-hours with known actual generation, 4,848 observations with positive planned generation formed the active supervised domain. The resulting chronological partition is summarized in Table 1. The forecast is defined at the local interval start. The previous-hour final measurement is assumed available after that interval ends,

and the plan is treated as issued before the forecast interval; the latter remains an explicit operational assumption because plan issue timestamps are absent from the archive.

Table 1. Audited active-hour modelling domain and chronological split

Indicator	Value
Raw hourly rows	28,296
Raw rows with known actual	22,320
Unique station-hours	11,736
Redundant source rows collapsed	16,560
Unique station-hours with known actual	9,792
Unique active station-hours	4,848
Solar power plants	3
Unique active dates	136
Train	3,038 rows / 81 dates (2022-04-01 to 2023-01-07)
Calibration	819 rows / 27 dates (2023-01-08 to 2023-02-16)
Held-out test	991 rows / 28 dates (2023-02-17 to 2023-03-23)

Forecasting formulation and leakage-control strategy

Let s denote a solar power plant and t the forecasted hourly interval. The target is measured active generation and the same-interval plan is an admissible predictor. Supervised modelling is restricted to unique station-hours with positive planned generation and known actual output. The final feature set includes planned generation and transforms, calendar/cyclic descriptors, station identity, actual-generation and plan lags at 1, 2, 3, 24, 48, and 72 hours, the one-hour historical plan-fact residual, and rolling statistics over 3, 6, 12, and 24 hours. Historical values are aligned to exact timestamps before active-hour filtering; preprocessing is fitted on training data only. Leakage controls prohibit the current actual value, current plan-fact deviation or target ratio, future targets, same-day finalized aggregates, and event variables without verifiable availability timestamps. Pipeline tests additionally confirm that changing the current target or future actual values does not alter earlier feature vectors. The full leakage-controlled workflow is shown in Figure 1.

Residual-Calibrated Neural Hybrid and comparator models

RCNH is a two-stage model. The primary branch is a Tabular MLP with 128 and 64 hidden units, ReLU activation and Adam optimization; its epoch count is selected only on a chronological inner-training tail. The residual calibrator is a HistGradientBoostingRegressor trained exclusively on calibration-period residuals and admissible prediction context.

The final prediction adds the estimated residual correction to the fixed MLP forecast and constrains generation to non-negative values. The held-out test targets are never used for preprocessing, base-model fitting, calibration, hyperparameter selection, or model selection. The two-stage RCNH architecture is illustrated in Figure 2.

Comparators are the plan schedule, one-hour persistence, Ridge, RandomForest, ExtraTrees, HistGradientBoosting, XGBoost, and the standalone Tabular MLP. Hyperparameters were fixed before held-out evaluation; the main experiment uses seed 42, with seeds 43 and 44 used only as sensitivity checks.

Experimental and statistical evaluation protocol

The primary experiment uses a chronological split: 3,038 training observations (through 7 January 2023), 819 calibration observations (8 January-16 February 2023), and 991 held-out test observations (17 February-23 March 2023). Performance is assessed with MAE, RMSE, R^2 , WAPE, and sMAPE, following established forecast-accuracy principles [18]. Although rank-based paired testing such as the Wilcoxon procedure is a classical nonparametric option for paired comparisons [19], the temporal dependence in this study motivated block-based inference.

Leakage-Controlled Pre-Modeling Pipeline for One-Hour-Ahead PV Forecasting

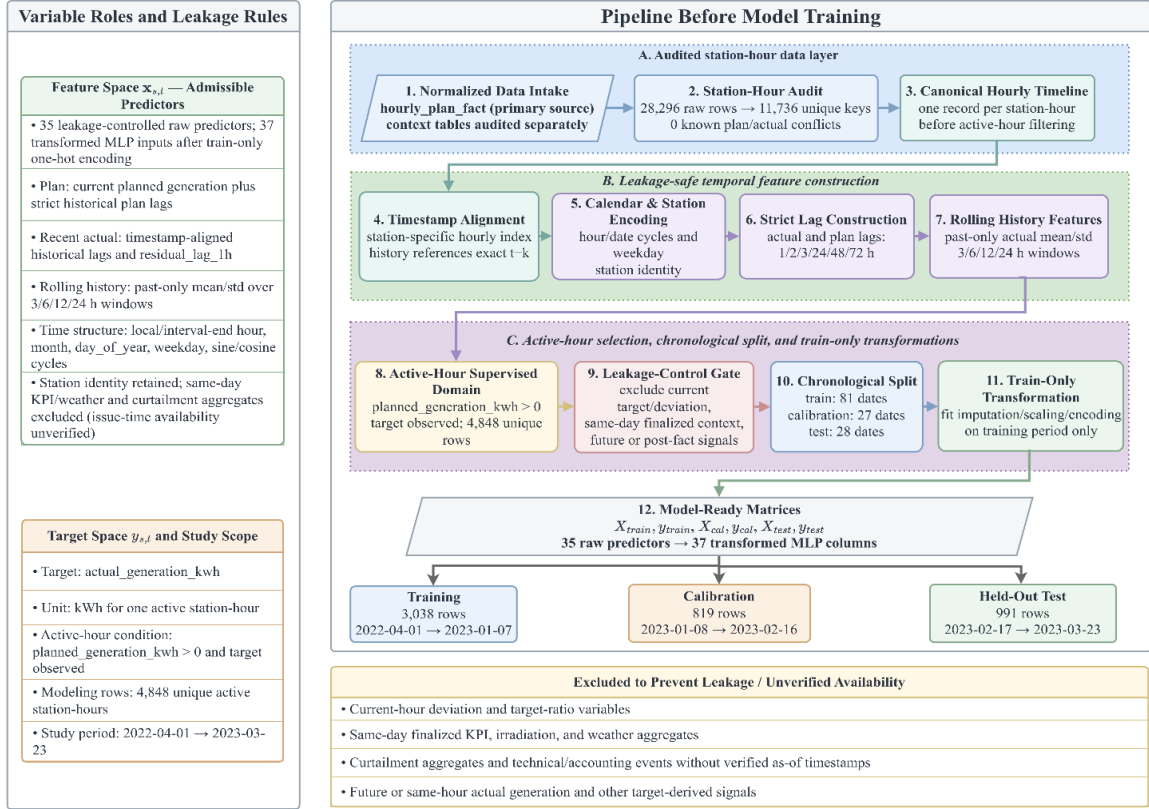


Figure 1. Leakage-controlled pre-modelling pipeline for one-hour-ahead PV generation forecasting

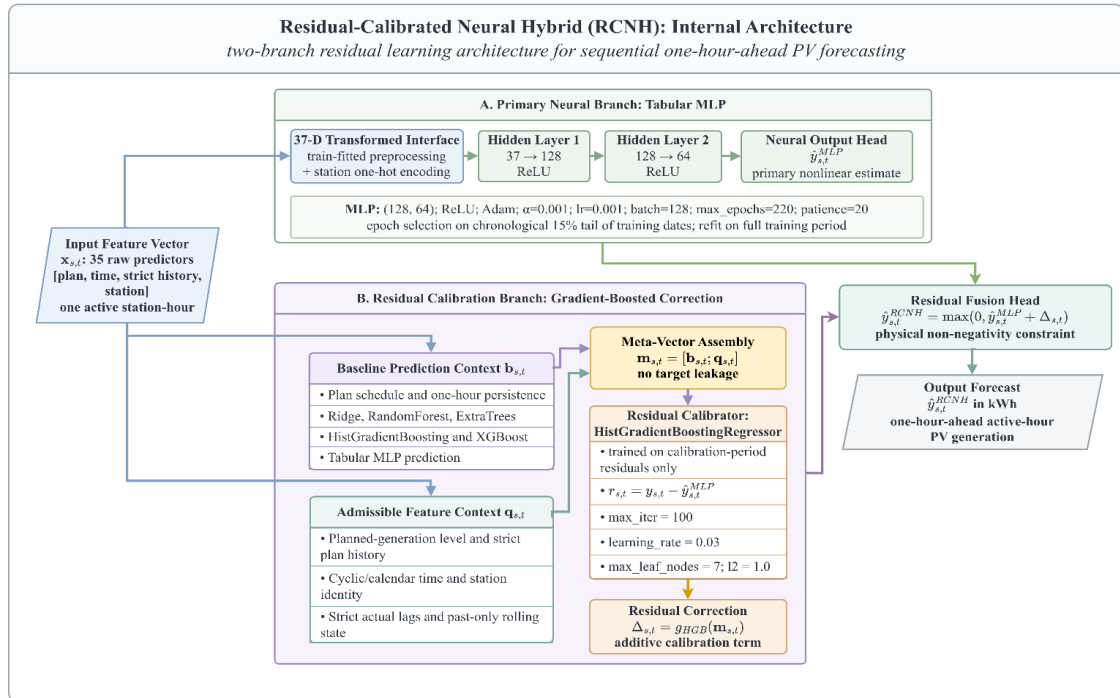


Figure 2. Architecture of the Residual-Calibrated Neural Hybrid forecasting model

Grouped permutation importance is used as model-specific interpretability evidence rather than as a causal ranking [20]. Uncertainty is quantified with a paired moving-day-block bootstrap using 2,000 replicates and three-day blocks; paired absolute-error differences are tested with a centered block-

bootstrap null distribution [21], and Holm correction is applied for multiple comparisons [22]. Robustness is examined through four chronological rolling-origin folds, calibration-window sensitivity, and a matched-label-budget diagnostic in which baselines are refitted on train plus calibration labels. Ablation experiments remove residual calibration, baseline-prediction context, lagged generation history, or plan-schedule features. Grouped permutation importance complements the ablation analysis by measuring sensitivity of the fitted model to perturbations of correlated feature groups [20].

Results of the study

Held-out performance and uncertainty

The comparative performance of all models on the held-out test period is presented in Table 2.

Table 2. Held-out test performance under the primary training/calibration protocol

Model	MAE, kWh	RMSE, kWh	R ²	WAPE, %	sMAPE, %
RCNH	2,591.80	3,906.53	0.908609	20.24	51.66
Ridge	2,802.96	4,054.10	0.901574	21.89	50.57
RandomForest	2,828.21	4,264.43	0.891096	22.08	43.43
Tabular MLP	2,923.98	4,273.60	0.890627	22.83	51.02
ExtraTrees	2,950.09	4,318.46	0.888319	23.03	45.39
XGBoost	2,885.25	4,345.78	0.886901	22.53	45.30
HistGradientBoosting	2,951.80	4,459.40	0.880910	23.05	44.21
One-hour persistence	3,954.09	6,362.27	0.757593	30.87	68.57
Plan schedule	5,460.12	7,900.31	0.626226	42.63	67.20

As shown in Table 2, on the held-out test period, RCNH achieved the lowest MAE and RMSE and the highest R² under the primary protocol: MAE = 2,591.80 kWh, RMSE = 3,906.53 kWh, and R² = 0.908609. Relative to the train-only Tabular MLP, MAE decreased by 11.36% and RMSE by 8.59%.

The three-day block-bootstrap 95% interval for RCNH MAE was 2,373-2,962 kWh and for RMSE 3,537-4,505 kWh. Against the train-only MLP, the relative MAE reduction was 11.36% (95% interval 6.98-14.95%), with Holm-adjusted two-sided p = 0.003998. The difference was also statistically supported against Ridge and the two operational baselines, whereas several close tree-based comparisons were not significant after multiplicity correction.

Matched-label-budget diagnostic

The results of the matched-label-budget diagnostic are presented in Table 3.

Table 3. Test performance when baseline models are refitted on train plus calibration labels

Model	MAE, kWh	RMSE, kWh	R ²
Tabular MLP	2,464.63	3,735.04	0.916457
ExtraTrees	2,659.36	3,877.07	0.909982
RCNH	2,591.80	3,906.53	0.908609
RandomForest	2,610.07	3,950.60	0.906535
XGBoost	2,646.84	3,970.03	0.905614
Ridge	2,729.29	3,982.07	0.905041
HistGradientBoosting	2,735.35	4,089.26	0.899859

Table 3 shows that the matched-label-budget diagnostic changes the interpretation of the primary result. When the standalone MLP is allowed to use the same 3,857 labeled observations available to the two-stage RCNH system, it achieves MAE = 2,464.63 kWh, RMSE = 3,735.04 kWh, and R² = 0.916457. Thus, part of RCNH's primary advantage reflects allocation of calibration labels rather than residual architecture alone.

Temporal robustness and sensitivity.

The numerical summary of the four rolling-origin validation folds is presented in Table 4.

Table 4. Numerical rolling-origin validation summary (four folds)

<i>Model</i>	<i>Mean MAE</i>	<i>SD MAE</i>	<i>Mean RMSE</i>	<i>SD RMSE</i>	<i>Mean R²</i>
<i>ExtraTrees</i>	2,072	371	3,065	537	0.868
<i>RandomForest</i>	2,016	414	3,099	645	0.864
<i>XGBoost</i>	2,076	366	3,122	537	0.861
<i>HistGradientBoosting</i>	2,227	396	3,311	553	0.848
<i>Plan schedule</i>	2,820	873	4,516	1,348	0.693
<i>One-hour persistence</i>	2,968	626	4,863	847	0.679
<i>RCNH</i>	4,843	3,076	6,107	2,823	0.424
<i>Tabular MLP</i>	5,409	3,619	6,498	3,515	0.355
<i>Ridge</i>	5,111	3,604	7,196	4,812	0.129

As summarized in Table 4, across four rolling-origin folds, ExtraTrees had the lowest mean RMSE (3,064.51 kWh), followed by RandomForest (3,098.58 kWh) and XGBoost (3,122.24 kWh), whereas RCNH reached 6,106.74 kWh with substantially larger fold-to-fold variation. Using only the most recent 50% or 75% of the calibration pool slightly improved average RCNH error relative to the full pool, reinforcing the dependence on calibration-window representativeness.

Ablation analysis

The effects of removing individual components of RCNH are presented in Table 5.

Table 5. RCNH ablation results on the final held-out test period

<i>Experiment</i>	<i>MAE, kWh</i>	<i>RMSE, kWh</i>	<i>R²</i>	<i>MAE change vs full</i>
<i>Full RCNH</i>	2,591.80	3,906.53	0.908609	+0.00
<i>Without residual calibration</i>	2,923.98	4,273.60	0.890627	+332.18
<i>Without baseline ensemble context</i>	2,585.35	3,901.76	0.908832	-6.45
<i>Without lagged generation history</i>	7,111.10	10,013.91	0.399479	+4,519.31
<i>Without plan schedule features</i>	2,599.85	3,826.49	0.912316	+8.05

Table 5 shows that removing residual calibration reproduced the standalone MLP and increased MAE by 332.18 kWh, while removing lagged generation history increased MAE to 7,111.10 kWh. By contrast, removing baseline-ensemble context or plan-schedule features produced little change after refitting, indicating strong redundancy among correlated predictors. Taken together, the results partially support the working hypothesis. Residual calibration improved the train-only MLP on the held-out test period, but the matched-label-budget and rolling-origin experiments showed that the improvement depends on calibration-window representativeness and temporal regime. Thus, the hypothesis was tested under both a fixed held-out split and additional temporal robustness checks, and the evidence supports a conditional rather than universal advantage of RCNH.

Discussion

The results support a bounded conclusion: residual calibration can materially improve a train-only neural predictor in a favorable held-out regime, but the gain is not an invariant property of the architecture. The matched-label-budget result and rolling-origin validation separate the effects of model design, access to later labeled data, and temporal regime. This distinction matters because the favorable final split should be interpreted as evidence of competitiveness rather than universal dominance. Tree ensembles were markedly more stable across temporal folds, while RCNH was more sensitive to calibration-period representativeness. The data audit is also central to interpretation:

28,296 exported rows correspond to 11,736 unique station-hours, of which only 4,848 form the active supervised sample. Reporting the true experimental unit prevents duplicated exports from artificially inflating sample size and apparent certainty. Feature analyses indicate that recent and rolling generation history is structurally important, but ablation and permutation results also reveal substantial predictor redundancy. The model can therefore be used as a forecasting component in a broader monitoring or digital-twin system, consistent with the role of digital-twin architectures described in prior work [8,9], but this study does not claim implementation of a complete digital twin.

These findings are consistent with prior photovoltaic-forecasting research showing that performance depends strongly on data characteristics, forecast horizon, and operating regime rather than on a single universally best model [3-7]. They also agree with broader evidence that tree-based ensembles remain strong benchmarks for medium-sized tabular prediction problems [13-17]. In this study, the additional contribution is to show that residual calibration can be useful on a favorable later regime while still losing temporal stability to tree ensembles when the regime changes.

The main limitations are the three-plant, 136-active-date scope; absence of hourly irradiance, cloud cover and temperature; unavailable issue timestamps for some operational indicators; and residual-calibrator sensitivity to temporal regime. Future work should extend validation across plants, seasons and years, incorporate verified meteorological and latency metadata, and examine online or drift-aware recalibration.

Conclusion

A leakage-controlled Residual-Calibrated Neural Hybrid was developed and evaluated for one-hour-ahead photovoltaic generation forecasting during active hours. The source audit reduced 28,296 raw rows to 11,736 unique station-hours and 4,848 active supervised observations, providing a defensible basis for chronological evaluation. Under the primary held-out protocol, RCNH achieved $MAE = 2,591.80$ kWh, $RMSE = 3,906.53$ kWh, and $R^2 = 0.908609$, with an 11.36% MAE reduction relative to the train-only MLP. However, a matched-label-budget MLP produced better final-test point metrics, and tree ensembles were more stable under rolling-origin validation. Therefore, RCNH should be considered a competitive regime-dependent residual-correction approach rather than a universally superior model. Its practical value is greatest when a recent, representative calibration window is available. The principal contribution is the combined evaluation protocol: unique station-hour auditing, strict prediction-time feature availability, chronological train-calibration-test separation, time-aware uncertainty quantification, matched information budgets, rolling-origin validation, and ablation analysis.

Data availability

The operational dataset contains industrial reporting and maintenance records and cannot be made publicly available because of confidentiality restrictions. The study reports the effective station-hour counts, chronological partitions, model specifications, uncertainty protocol, and aggregate results required to understand the experimental design. Raw source duplication was handled by deterministic station-hour consolidation of identical known values; conflicting known values would terminate the analytical pipeline.

Acknowledgment

This research was funded by the Ministry of Science and Higher Education of the Republic of Kazakhstan, Grant No. AP26105051.

References

[1] Inman R.H., Pedro H.T.C., Coimbra C.F.M. Solar forecasting methods for renewable energy integration // *Progress in Energy and Combustion Science*. 2013. Vol. 39, No. 6. P. 535-576. DOI: <https://doi.org/10.1016/j.pecs.2013.06.002>

- [2] Diagne M., David M., Lauret P., Boland J., Schmutz N. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids // *Renewable and Sustainable Energy Reviews*. 2013. Vol. 27. P. 65-76. DOI: <https://doi.org/10.1016/j.rser.2013.06.042>
- [3] Antonanzas J., Osorio N., Escobar R., Urraca R., Martinez-de-Pison F.J., Antonanzas-Torres F. Review of photovoltaic power forecasting // *Solar Energy*. 2016. Vol. 136. P. 78-111. DOI: <https://doi.org/10.1016/j.solener.2016.06.069>
- [4] Voyant C., Notton G., Kalogirou S., Nivet M.L., Paoli C., Motte F., Fouilloy A. Machine learning methods for solar radiation forecasting: A review // *Renewable Energy*. 2017. Vol. 105. P. 569-582. DOI: <https://doi.org/10.1016/j.renene.2016.12.095>
- [5] Ahmed R., Sreeram V., Mishra Y., Arif M.D. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization // *Renewable and Sustainable Energy Reviews*. 2020. Vol. 124. Article 109792. DOI: <https://doi.org/10.1016/j.rser.2020.109792>
- [6] Iheanetu K.J. Solar photovoltaic power forecasting: A review // *Sustainability*. 2022. Vol. 14, No. 24. Article 17005. DOI: <https://doi.org/10.3390/su142417005>
- [7] Gaboitaolelwe J., Zungeru A.M., Yahya A., Lebekwe C.K., Vinod D.N., Salau A.O. Machine learning based solar photovoltaic power forecasting: A review and comparison // *IEEE Access*. 2023. Vol. 11. P. 40820-40845. DOI: <https://doi.org/10.1109/ACCESS.2023.3270041>
- [8] Fuller A., Fan Z., Day C., Barlow C. Digital twin: Enabling technologies, challenges and open research // *IEEE Access*. 2020. Vol. 8. P. 108952-108971. DOI: <https://doi.org/10.1109/ACCESS.2020.2998358>
- [9] Angelova D.D. et al. A review on digital twins and its application in the modeling of photovoltaic installations // *Energies*. 2024. Vol. 17, No. 5. Article 1227. DOI: <https://doi.org/10.3390/en17051227>
- [10] Kaufman S., Rosset S., Perlich C., Stitelman O. Leakage in data mining: Formulation, detection, and avoidance // *ACM Transactions on Knowledge Discovery from Data*. 2012. Vol. 6, No. 4. Article 15. DOI: <https://doi.org/10.1145/2382577.2382579>
- [11] Hyndman R.J., Athanasopoulos G. *Forecasting: Principles and Practice*. 3rd ed. Melbourne: OTexts, 2021. URL: <https://otexts.com/fpp3/>
- [12] Bergmeir C., Hyndman R.J., Koo B. A note on the validity of cross-validation for evaluating autoregressive time series prediction // *Computational Statistics & Data Analysis*. 2018. Vol. 120. P. 70-83. DOI: <https://doi.org/10.1016/j.csda.2017.11.003>
- [13] Friedman J.H. Greedy function approximation: A gradient boosting machine // *The Annals of Statistics*. 2001. Vol. 29, No. 5. P. 1189-1232. DOI: <https://doi.org/10.1214/aos/1013203451>
- [14] Breiman L. Random forests // *Machine Learning*. 2001. Vol. 45. P. 5-32. DOI: <https://doi.org/10.1023/A:1010933404324>
- [15] Chen T., Guestrin C. XGBoost: A scalable tree boosting system // *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016. P. 785-794. DOI: <https://doi.org/10.1145/2939672.2939785>
- [16] Wolpert D.H. Stacked generalization // *Neural Networks*. 1992. Vol. 5, No. 2. P. 241-259. DOI: [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)
- [17] Grinsztajn L., Oyallon E., Varoquaux G. Why do tree-based models still outperform deep learning on typical tabular data? // *Advances in Neural Information Processing Systems*. 2022. Vol. 35. P. 507-520. DOI: <https://doi.org/10.52202/068431-0037>
- [18] Hyndman R.J., Koehler A.B. Another look at measures of forecast accuracy // *International Journal of Forecasting*. 2006. Vol. 22, No. 4. P. 679-688. DOI: <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- [19] Wilcoxon F. Individual comparisons by ranking methods // *Biometrics Bulletin*. 1945. Vol. 1, No. 6. P. 80-83. DOI: <https://doi.org/10.2307/3001968>
- [20] Fisher A., Rudin C., Dominici F. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously // *Journal of Machine Learning Research*. 2019. Vol. 20, No. 177. P. 1-81. URL: <https://www.jmlr.org/papers/v20/18-760.html>
- [21] Kunsch H.R. The jackknife and the bootstrap for general stationary observations // *The Annals of Statistics*. 1989. Vol. 17, No. 3. P. 1217-1241. DOI: <https://doi.org/10.1214/aos/1176347265>
- [22] Holm S. A simple sequentially rejective multiple test procedure // *Scandinavian Journal of Statistics*. 1979. Vol. 6, No. 2. P. 65-70. DOI: <https://doi.org/10.2307/4615733>