

An Oracle Persistence Framework Using Capacity Information to Diagnose Concept Drift in Long Term Photovoltaic Power Forecasting

changxuan cao

East China University of Technology

zhifeng liu

20232119040@ecut.edu.cn

East China University of Technology

jiaijun xu

East China University of Technology

xueqiang deng

East China University of Technology

Research Article

Keywords: PV power forecasting, Concept drift, Capacity degradation, Oracle-Persistence diagnostic, System-state monitoring, Dual-site validation, BiLSTM-Attention

Posted Date: September 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10866555/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

An Oracle Persistence Framework Using Capacity Information to Diagnose Concept Drift in Long Term Photovoltaic Power Forecasting

Changxuan Cao¹ · Zhifeng Liu^{1*} · Jiajun Xu² · Xueqiang Deng²

Abstract

Reliable photovoltaic (PV) power forecasts support grid operation, yet models trained on historical data can silently lose accuracy when an installation's effective capacity changes through degradation, faults, or maintenance. Most forecasting studies implicitly treat the PV system as time-invariant and therefore cannot tell whether a loss of cross-period accuracy originates in the weather or in the system itself. We address this with a capacity-aware diagnostic, evaluated on long-term records from the Desert Knowledge Australia Solar Centre (DKASC, Australia) and PVDAQ Site 34 (Colorado, USA) — 26 combined site-years.

PV output is represented as $Y(t) = C(t) \cdot g(X(t)) + \epsilon(t)$, separating a latent effective-capacity factor $C(t)$ from the weather-to-power mapping $g(X(t))$. We introduce the Oracle-Persistence diagnostic, which contrasts persistence, cross-period transfer (Oracle-Cross), and within-period predictability (Oracle-Self) to assess whether transfer failure is consistent with effective-capacity change. In this two-site study, the diagnostic produced contrasting patterns for a severely drifting system and a stable site. At DKASC, cross-period R^2 was -0.251 with weather features alone; adding a retrospective capacity feature raised it to 0.955 (a diagnostic upper bound), while a strictly pre-origin, leakage-free capacity estimate still reached $R^2 = 0.873$, far above persistence (0.646). This recovery supports effective-capacity change as an important contributor to transfer failure, but does not by itself identify the physical mechanism. At PVDAQ Site 34, cross-period R^2 was 0.903 with stable capacity indicators. Within a separate common configuration pipeline, robust scaling produced the only statistically supported gain over MinMax scaling (nominal Diebold–Mariano $p < 0.001$), and the best configuration reached $R^2 = 0.862$ on a deliberately non-stationary test set. Code and processed data are available from the

corresponding author upon reasonable request. The Oracle-Persistence diagnostic is therefore presented as a candidate low-cost monitoring tool for detecting when a deployed forecaster no longer tracks changes in its own hardware.

Keywords: PV power forecasting; Concept drift; Capacity degradation; Oracle-Persistence diagnostic; System-state monitoring; Dual-site validation; BiLSTM-Attention

* *Zhifeng Liu*

2023211904@ecut.edu.cn

Changxuan Cao

1028359003@qq.com

Jiajun Xu

15579178689@163.com

Xueqiang Deng

2023211863@ecut.edu.cn

ORCID IDs:

Changxuan Cao: <https://orcid.org/0009-0005-7143-4628>

Zhifeng Liu (Corresponding Author): <https://orcid.org/0000-0001-6524-0038>

¹ School of Artificial Intelligence and Information Engineering, East China University of Technology, Nanchang, China

² School of Nuclear Science and Engineering, East China University of Technology, Nanchang, China

1. Introduction

1.1 Background and Motivation

Global installed solar photovoltaic (PV) capacity surpassed 1.6 TW by the end of 2023 [10]. As PV penetration rises, accurate short-term power forecasting — predicting output minutes to hours ahead — has become a critical enabling technology for grid dispatch, battery scheduling, and electricity-market participation [19,20,25]. Yet despite the maturity of forecasting methods [19,20,21,25,51], almost all published studies share one implicit and physically unjustifiable assumption: that the PV system is a time-invariant converter of solar radiation into electrical power.

Real installations, however, undergo progressive module degradation, inverter aging, soiling accumulation, connector corrosion, and discrete events such as sub-array disconnection [16,17,18,33,60,62]. These processes steadily alter the effective capacity $C(t)$, introducing non-stationarity that violates the foundational assumption of most forecasting models. Ignoring it produces

systematic bias, inflated error at high irradiance, and — most damagingly — negative transfer, in which knowledge learned from the training set actively degrades test-set predictions.

1.2 Research Gaps

Three practical gaps motivate this study. First, most PV forecasting evaluations span relatively short periods and thus offer limited evidence on performance after major changes in effective capacity [19–21]. Second, generic concept-drift detectors flag distributional change but do not indicate whether a loss of transferability stems from system state or from meteorological variability [1,14,22,23,47,48]. Third, adaptation strategies are often evaluated on different datasets and model configurations, which obscures their relative value. Recent energy-domain work further develops proactive drift adaptation and drift-aware data augmentation for online forecasting [58,59].

Recent deep-learning studies demonstrate strong short-term PV forecasting performance — for example, CNN-LSTM forecasting with Gaussian-mixture error correction [53], weather-type clustering combined with a CNN-BiLSTM hybrid [54], a phase-space-reconstruction GAT-LSTM model [55], and further hybrids such as CFOA-CNN-BiLSTM-Attention [56] and TCN-ECANet-GRU [57] — but their evaluation settings address short-horizon accuracy on comparatively stable records and do not directly test long-term changes in effective system capacity. Our study differs in temporal scope and objective: rather than competing on short-term accuracy, it evaluates cross-period transfer across two long records and asks whether an explicit estimate of effective capacity provides evidence consistent with the observed transfer failure. This comparison should therefore not be read as a claim that the present architecture outperforms these methods, because datasets and protocols differ.

1.3 Principal Contributions

This study makes three contributions:

- (1) Theory and diagnostic. We formalize the link between unmodeled effective-capacity variation and systematic forecast degradation. The resulting Oracle-Persistence quantities convert the question of whether a loss is consistent with weather variation or system-state change into an operational consistency check, rather than a causal identification test.
- (2) Dual-site contrast validation. We evaluate the framework on two geographically and operationally distinct PV plants: the Danish DKASC site, which exhibits strong cross-period degradation, and the

U.S. PVDAQ site, which is comparatively stable. The contrasting diagnostic patterns provide feasibility evidence for the proposed interpretation while not establishing universality across PV fleets.

(3) Operational availability and reproducibility. We provide a strictly pre-origin (leakage-free) capacity-feature experiment showing that capacity information improves cross-period transferability when no observations from the forecast target interval are used (Oracle-Cross $R^2 = 0.873$, versus a 0.955 retrospective upper bound and -0.251 without capacity). Within a separate common configuration pipeline, robust scaling is the only change with a statistically supported improvement (nominal Diebold–Mariano $p < 0.001$). The framework and processed data are available from the corresponding author upon reasonable request. The six candidate adaptation strategies are reported as supporting experimental analyses in Section 5.5 rather than as a complete controlled ranking.

2. Theoretical Framework

2.1 Degradation Model and Drift Taxonomy

The standard static regression $Y(t) = g(X(t)) + \varepsilon(t)$ treats the system as time-invariant. We instead propose a multiplicative structure:

$$Y(t) = C(t) \cdot g(X(t)) + \varepsilon(t) \quad (\text{Equation 1})$$

Information-availability and proxy constraints. In the retrospective analyses, $\hat{C}(t)$ was estimated from a trailing 30-day window of observed power using a rolling P95 statistic. For operational forecasting, this window must end strictly before the forecast origin; using observations from the forecast target interval would constitute target leakage. The retrospective capacity-feature Oracle-Cross result is therefore reported only as a diagnostic upper bound. Moreover, rolling P95 is a proxy rather than a direct capacity measurement: it can also respond to irradiance regime, cloudiness, clipping, curtailment, and missing-data patterns. A performance recovery after adding $\hat{C}(t)$ is therefore interpreted as capacity-consistent evidence, not as identification of a unique physical mechanism.

where $C(t) \in (0,1]$ is the effective capacity factor at time t , $g(X(t))$ is the ideal weather-to-power mapping, and $\varepsilon(t)$ is noise. In practice $C(t)$ is latent and must be estimated; here we use the rolling 30-day P95 estimator described in Section 4.1 (Strategy S4). Long-term non-stationarity arises when $C(t)$ varies while $X(t)$ remains statistically stable. The formulation accommodates all three classical drift types — gradual (monotone $C(t)$), sudden (step $C(t)$), and recurring (periodic $C(t)$) — plus a fourth

recovery type identified in this study. Because each physical mechanism leaves a distinct statistical signature, we formalise them as four drift types in Table 1 (Section 2.1).

Gradual (P1)	Component aging, soiling	C(t) monotone decline	Segment R² declining over time
Sudden (P2)	Sub-array disconnect, event	C(t) step change	Pettitt test + Oracle-Cross collapse
Recurring (P3)	Seasonal temperature/soiling	C(t) periodic oscillation	Monthly residual seasonality
Candidate recovery	Repair, cleaning, reconnection	C(t) step increase	Positive capacity step; mechanism requires records

Table 1. Working taxonomy of effective-capacity drift patterns. Recovery is treated as a candidate pattern because the August 2025 event lacks maintenance records.

The four drift types are formally defined as follows. P1 (Gradual drift): C(t) follows a monotone decline due to continuous physical aging — module photodegradation, soiling accumulation, or connector corrosion. Statistical signature: segment R² declines progressively over time. P2 (Sudden drift): C(t) undergoes a discrete step change caused by a specific event — sub-array disconnection, inverter failure, or partial shading. Statistical signature: Pettitt change-point test significant; Oracle-Cross R² collapses at the event boundary. P3 (Recurring drift): C(t) exhibits periodic oscillation driven by seasonal temperature coefficients, dust cycles, or monsoonal cloud patterns. Statistical signature: monthly residuals show systematic seasonality. Recovery (candidate pattern, this study): C(t) undergoes a positive step change following maintenance intervention — reconnection, cleaning, or component replacement. Statistical signature: R² rises post-event; P95 power increases without corresponding GHI change.

2.2 Oracle-Persistence Diagnostic Framework

Three model-dependent reference scores define the diagnostic space: (1) R²_{pers} (persistence baseline) measures temporal autocorrelation and provides a practical benchmark, not a universal lower bound. (2) R²_{oc} (Oracle-Cross) is obtained by training the selected model and feature pipeline on historical data and evaluating it on later data; it measures cross-period transfer for that pipeline rather than the best performance achievable by any model. (3) R²_{os} (Oracle-Self) is obtained using data drawn from the evaluation period and serves as a model- and feature-dependent within-period reference. It is not an information-theoretic predictability ceiling and should not be compared as a universal upper bound across different feature sets.

Under approximate stationarity, matched feature definitions, and comparable evaluation protocols, one

expects $R^2_{\text{pers}} \leq R^2_{\text{model}}$ and R^2_{oc} to approach R^2_{os} . This is a diagnostic expectation rather than a mathematical ordering; model misspecification, finite samples, or different feature sets can violate it. Three corollaries follow:

Corollary 1 (Drift warning): $R^2_{\text{oc}} < R^2_{\text{pers}}$ indicates that the evaluated historical mapping transfers less effectively than persistence. This pattern is consistent with drift, but can also arise from model misspecification or inadequate features.

Corollary 2 (Severe transfer failure): $R^2_{\text{oc}} < 0$ indicates that the cross-period model performs worse than a constant-mean reference. Persistence should be preferred only when its measured performance also exceeds R^2_{oc} .

Corollary 3 (Capacity-consistency test): a substantial increase in R^2_{oc} after adding a strictly pre-origin estimate of $C(t)$ supports an association consistent with effective-capacity change contributing to transfer failure. It does not establish causal attribution or identify a physical fault mechanism.

2.3 NWP Marginal Value (Equation 2)

The potential marginal value of future weather information depends on the conditional uncertainty of future irradiance:

$$I_{\text{NWP}} = H(\text{GHI}_{\{t+h\}} | \text{GHI}_t) - H(\text{GHI}_{\{t+h\}} | \text{GHI}_t, \text{NWP}_{\{t+h\}}) \quad (\text{Equation 2})$$

When short-horizon irradiance is strongly persistent, additional weather forecasts may add little information and can even introduce calibration error; at more variable sites, high-quality forecasts may carry greater marginal value. Section 5.8 evaluates only the direction of this hypothesis.

3. Study Sites and Data

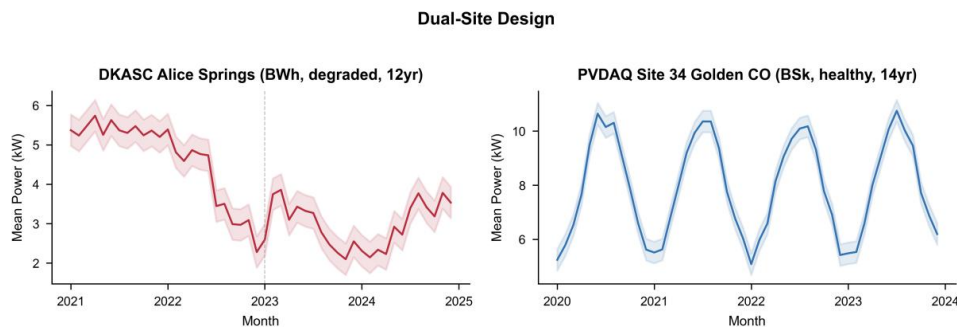


Fig. 1. Dual-site design: DKASC Alice Springs (BWh, degraded, 12 yr) versus PVDAQ Site 34 Golden CO (BSk, comparatively stable, 14 yr).

3.1 Primary Site: DKASC Alice Springs, Australia

The Desert Knowledge Australia Solar Centre (DKASC) [3,52] is located at Alice Springs, Northern Territory, Australia (23.762°S, 133.875°E, 545 m). Its BWh hot-desert climate delivers an annual mean GHI of ~2,200 kWh/m² and more than 300 clear-sky days per year. The site has been monitored since 2013; this study analyses the most recent 4-year subset (2021–2025; 187,647 records at 5-minute resolution), split chronologically 70/10/20 into training, validation, and test sets.

The 4-year subset (2021–2025) was chosen for two reasons. First, pre-2021 records reflect a different instrumentation configuration at DKASC, in which sensor recalibrations introduce a step-level discontinuity that confounds degradation attribution; the 4-year window post-dates these changes and provides a self-consistent measurement baseline. Second, this window captures the full sequence of interest — the pre-break healthy phase (2021–2022), the structural break (March 2023), the severely degraded phase (2023–2024), and the subsequent capacity-recovery event (August 2025) — thereby maximising diagnostic coverage of the four drift types in Table 1. The long-term degradation-rate estimation in Section 5.1 is based on the same 2021–2025 subset (49 monthly η_{eff} observations).

3.2 Stable-Site Contrast: PVDAQ Site 34, Golden, Colorado, USA

PVDAQ Site 34 (39.74°N, 105.17°W, 1,829 m) operates under a BSk cool semi-arid climate, with ~2,450 annual sunshine hours, pronounced seasonal variation (including winter snowfall), and moderate summer cloud cover. The dataset spans October 2010 to December 2023 (14 years; nameplate capacity ~13.4 kW; 15-minute resolution) and was obtained from the NREL Open Energy Data Initiative (OEDI) under a Creative Commons Attribution licence (DOI: 10.25984/1846021).

A unit-conversion correction was applied to the raw `ac_power_hw` field: values before August 2018 were divided by 100 to obtain kW, and later values by 10,000. Cross-year peak-power consistency served as a diagnostic check. PVDAQ Site 34 was selected for climatic contrast; its apparent stability emerged only after the diagnostic analysis was applied.

Property	DKASC Alice Springs	PVDAQ Site 34 Golden CO
Location	23.8°S, 133.9°E	39.7°N, 105.2°W
Köppen climate	BWh (hot desert)	BSk (cool semi-arid)
Elevation	545 m	1,829 m
Data span	Aug 2013 – Aug 2025 (12 yr)	Oct 2010 – Dec 2023 (14 yr)
Resolution	5 minutes	15 minutes
Records analysed	187,647 (4-yr subset)	369,705 (full span)

Property	DKASC Alice Springs	PVDAQ Site 34 Golden CO
Degradation rate	-17.03%/yr ($p < 0.0001$)	-0.13%/yr ($p = 0.35$, n.s.)
Pettitt test	$p = 1.4 \times 10^{-7}$ (break Mar 2023)	$p = 0.41$ (no break)
Role in study	Primary: drift & forecasting	Stable-site contrast

Table 2. Site characteristics and long-term diagnostic summaries.

4. Methodology

4.1 Six Adaptive Strategies

Strategy	Hypothesis	Key Implementation	Target Component
S0: Baseline tuning	Better hyperparameters improve mapping	RobustScaler + 4h window + 120 epochs	Optimise $f \approx C \cdot g$
S1: Shortened window	Recent data reduces drift amplitude	2-year training subset	Reduce $\Delta C(\text{train/test})$
S2: Capacity normalisation	Normalising by $\hat{C}(t)$ isolates $g(X)$	Predict $CF = Y/\hat{C}(t)$; inverse-transform	Remove $C(t)$
S3: Anomaly cleaning	Remove corrupted samples	Conditional quantile filter (P35)	Reduce $\varepsilon(t)$
S4: Capacity-aware features	Explicit $\hat{C}(t)$ resolves latent variable	Rolling P95 + capacity ratio as features	$C(t)$: latent $\rightarrow$ observed
S5: NWP integration	Future weather reduces uncertainty	Open-Meteo 1h-ahead + 3-day forecasts	Improve $g(X_{\{t+h\}})$

Table 3. Six candidate adaptation strategies and their intended targets.

4.2 Model Architecture

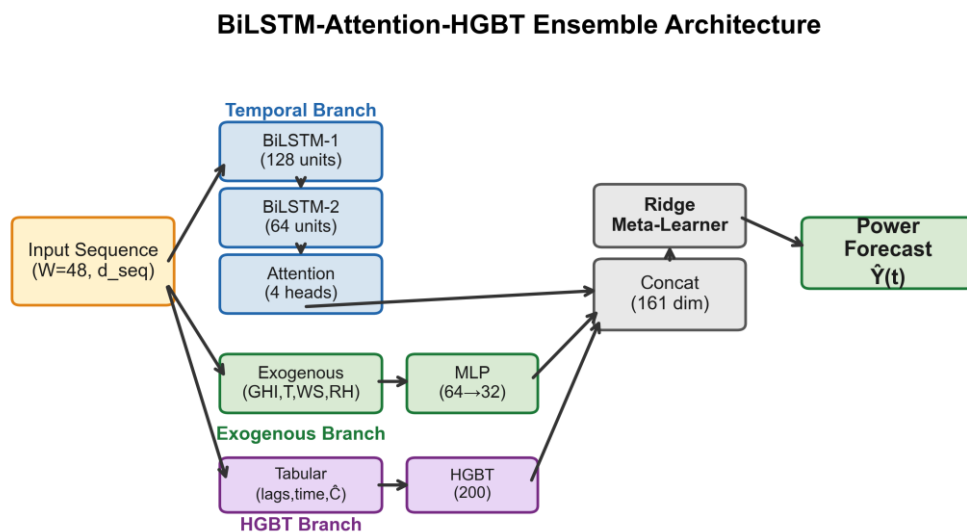


Fig. 2. BiLSTM-Attention-HGBT architecture. Three branches (temporal BiLSTM+Attention, exogenous weather, HGBT tabular) fused by a Ridge meta-learner stacking ensemble.

The forecasting model combines three parallel branches fused by a Ridge meta-learner. (1) BiLSTM-Attention branch: the input sequence (window length W , feature dimension d_{seq}) is processed by two stacked bidirectional LSTM layers [11] (128 and 64 units; effective outputs 256 and 128 after concatenation), followed by 4-head multi-head attention [12,28] (key dimension 32 per head); the branch outputs a 128-dimensional temporal representation. (2) Exogenous branch: current-timestep meteorological features (GHI, temperature, wind speed, humidity; dimension $d_{exo} = 4$ or 6 with NWP) are passed through a two-layer MLP (64→32 units, ReLU activation), producing a 32-dimensional vector. (3) HGBT branch [29]: all tabular features (lag values, time-of-day encodings, capacity features) are processed by HistGradientBoosting (max_iter=200), producing a scalar prediction. The Ridge meta-learner ($\alpha=1.0$) receives the concatenation of all three branch outputs [$128 + 32 + 1 = 161$ dimensions] and produces the final power forecast. S0 training configuration: RobustScaler preprocessing, 4-hour input window (48 steps at 5-minute resolution), Huber loss ($\delta=1.0$), Adam [37] ($lr=8\times 10^{-4}$, clipnorm=1.0), early stopping (patience=12), max 120 epochs.

4.3 Evaluation Protocol and Interpretive Scope

The R^2 values reported in Sections 5.2 and 5.5 serve different evaluation roles and should not be interpreted as a single performance ranking. The S0 result ($R^2 = 0.862$) summarises configuration performance within the common chronologically split forecasting pipeline. The Oracle-Cross values at DKASC (-0.251 , 0.955 , and 0.873) quantify cross-period transfer under weather-only, retrospective capacity, and strictly pre-origin capacity information, respectively. The PVDAQ value ($R^2 = 0.903$) is a separate stable-site transfer contrast.

Oracle-Self is used only as a model- and feature-dependent within-period reference. It indicates how predictable a period appears when model development and evaluation draw from the same temporal regime; it is not a universal ceiling across different feature sets. Likewise, Oracle-Cross denotes the transfer performance of the selected model and feature pipeline, not the best possible cross-period performance in an absolute sense.

For the operationally available capacity-feature experiment, every rolling P95 window ends strictly before the forecast origin, and no observations from the forecast target interval are used. The retrospective result ($R^2 = 0.955$) relaxes this information boundary and is therefore reported only as a diagnostic upper bound. Because leakage risk depends on timestamp alignment, forecast horizon, and data latency, implementations should document these quantities explicitly when reproducing or

deploying the diagnostic.

4.4 Statistical Testing and Inferential Scope

Four statistical procedures were used: (1) the Diebold–Mariano (DM) test [6] for pairwise comparison of forecast errors; (2) bootstrap confidence intervals [15] with 1,000 resamples; (3) the Pettitt non-parametric change-point test [5] applied to monthly effective-conversion-efficiency estimates; and (4) the Kolmogorov–Smirnov test for changes in the marginal power distribution. The five-minute and monthly records are temporally dependent, so conventional resampling and null distributions may be anti-conservative when that dependence is not explicitly modelled. Accordingly, the reported p values are treated as nominal and descriptive; effect sizes, confidence intervals, temporal patterns, and the cross-site contrast carry greater interpretive weight. Confirmatory evaluation would require blocked resampling, dependence-robust DM variance estimation, and autocorrelation-aware change-point calibration.

5. Results

5.1 DKASC Degradation Characterisation

Mean daytime power fell from 5.84 kW (2021) to 2.48 kW (2024) — a cumulative decline of 57.5% — with no corresponding reduction in GHI (~445–460 W/m² throughout). The Pettitt test on monthly η_{eff} (effective conversion efficiency = mean daytime power / mean daytime GHI, in kW per W/m²) identifies March 2023 as a structural break ($U = 524.0$, $p = 1.41 \times 10^{-7}$). The degradation rate inferred from the 49 monthly η_{eff} observations over 2021–2025 is $-17.03\%/ \text{year}$ ($p < 0.0001$), an order of magnitude beyond typical crystalline-silicon aging [2]. We stress that $-17.03\%/ \text{yr}$ is the net compound rate of this specific installation — combining irreversible aging with discrete, recoverable fault events — and is presented as evidence of compound, fault-driven drift rather than as a generalisable PV degradation rate.

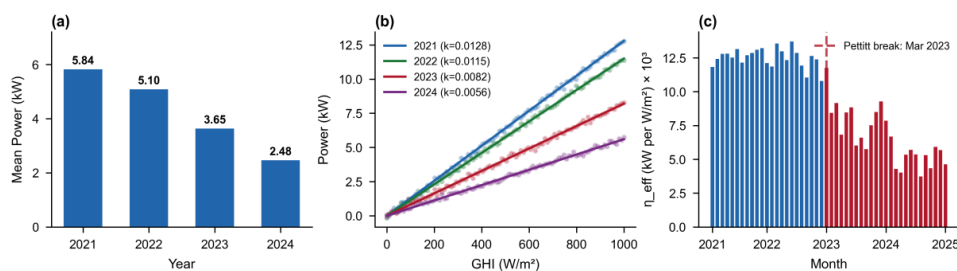


Fig. 3. DKASC system degradation: (a) annual power decline 5.84→2.48 kW (−57.5%); (b) GHI–

Power slope shift by year; (c) monthly η_{eff} with Pettitt change point (March 2023, $p = 1.41 \times 10^{-7}$).

Year	Mean Power (kW)	P95 (kW)	Mean GHI (W/m ²)	η_{eff} (kW per W/m ²)	Δ vs 2021
2021	5.84	7.79	~460	0.00587	Baseline
2022	5.29	7.63	~455	0.00611	−9.4%
2023	3.08	7.23	~450	0.00446	−47.3%
2024	2.48	3.84	~445	0.00331	−57.5%

Table 4. DKASC year-by-year statistics showing progressive capacity collapse.

Physical interpretation of the anomalous degradation rate. The $-17.03\%/year$ compound rate is $17\text{--}34\times$ the $0.5\text{--}1.0\%/year$ typical of normal crystalline-silicon aging [2,16,17,61], pointing to compound degradation — progressive physical deterioration combined with discrete, recoverable fault events. Three lines of evidence support this. First, the Pettitt structural break in March 2023 ($p = 1.4 \times 10^{-7}$) coincides with a step-level capacity collapse (P95: $7.23 \rightarrow 3.84$ kW) consistent with an abrupt system-level change rather than smooth aging, but not specific to an electrical fault. Second, the August 2025 recovery (mean power $2.29 \rightarrow 3.44$ kW) shows a recoverable change consistent with cleaning, reconnection, or component replacement, whereas irreversible photodegradation cannot recover without panel replacement. Third, the BWh hot-desert environment — with extreme diurnal temperature cycling ($\Delta T > 30$ °C/day), high UV intensity, and dust storms — accelerates connector corrosion and thermal fatigue [18,60]. The $-17.03\%/year$ figure thus represents the net compound degradation rate of this specific installation, encompassing both irreversible aging and fault-induced losses. This does not diminish its diagnostic value: such severe drift provides a demanding stress-test for the Oracle-Persistence framework, and the recovery event extends the drift taxonomy.

5.2 Oracle-Persistence Diagnostics (DKASC)

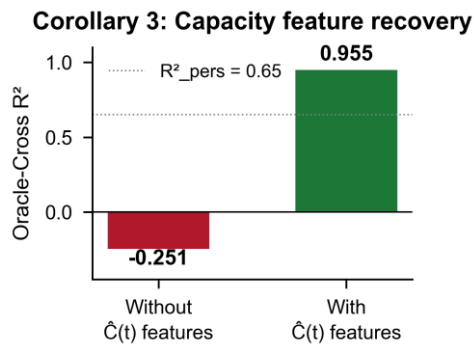


Fig. 4. DKASC cross-period diagnostic. Adding retrospective capacity features raised Oracle-Cross R^2 from -0.251 to 0.955 . This is capacity-consistent but may reflect target leakage, proxy confounding, or model misspecification.

With weather features, $R^2_{oc} = -0.251$ was below $R^2_{pers} = 0.650$ and below zero, triggering the two transfer-failure warnings. Adding a retrospective capacity feature increased R^2_{oc} to 0.955; because this estimator can draw on information up to the forecast target, we report it only as a diagnostic upper bound. Under a strictly pre-origin, leakage-free capacity estimate whose window ends before the forecast origin (Section 2.1), R^2_{oc} was 0.873 — below the 0.955 upper bound but far above both the -0.251 weather-only baseline and persistence ($R^2_{pers} = 0.650$). The $0.955 \rightarrow 0.873$ gap describes the value of concurrent versus pre-origin capacity information. The 0.873 result is consistent with effective-capacity change being an important contributor to transfer failure, but it does not identify a physical mechanism or constitute prospective deployment validation. The marginal power distributions also differed substantially (KS D = 0.324); given the large, serially dependent sample, the nominal $p \approx 0$ is not interpreted as independent evidence.

Configuration	R^2_{pers}	R^2_{oc}	R^2_{os}	KS D	KS p	Diagnosis
Raw weather features	0.650	-0.251	0.379	0.324	≈ 0	Cor. 1 & 2 triggered
+Capacity features $\hat{C}(t)$	0.650	0.955	—	—	—	Capacity-consistent interpretation supported

Table 5. Oracle-Persistence diagnostic results for DKASC.

5.3 Assessment of the Proposed Drift Patterns

P1 (Gradual drift): Test-set R^2 increased from 0.809 in 2024 H2 to 0.907 in 2025 H2, contrary to the proposed signature of progressively declining accuracy. The observed pattern therefore does not confirm P1 during the test period. Adaptation through recent-power lag features is one possible explanation, but requires prospective evaluation.

Candidate recovery pattern. Mean power rose from 2.29 kW in 2025 H1 to 3.44 kW in August 2025. This positive step is consistent with cleaning, reconnection, or component replacement, but maintenance records would be needed to establish the mechanism. We therefore describe it as an observed recovery event consistent with maintenance intervention, rather than proposing it as a confirmed new drift type.

P2 (Sudden drift). Three observations support an abrupt-change interpretation: a deterioration in cross-period transfer around 2023, a Pettitt change point in March 2023 (nominal $p = 1.4 \times 10^{-7}$), and a P95

decrease from 7.23 to 3.84 kW. These are consistent with, but do not by themselves identify, a discrete system event.

P3 (Recurring drift). No clear recurring residual pattern was observed at DKASC. PVDAQ Site 34 did show a seasonal residual pattern (Section 5.7), although seasonality alone does not establish recurring capacity drift.

5.4 Descriptive Partition of Diagnostic R^2 Gaps

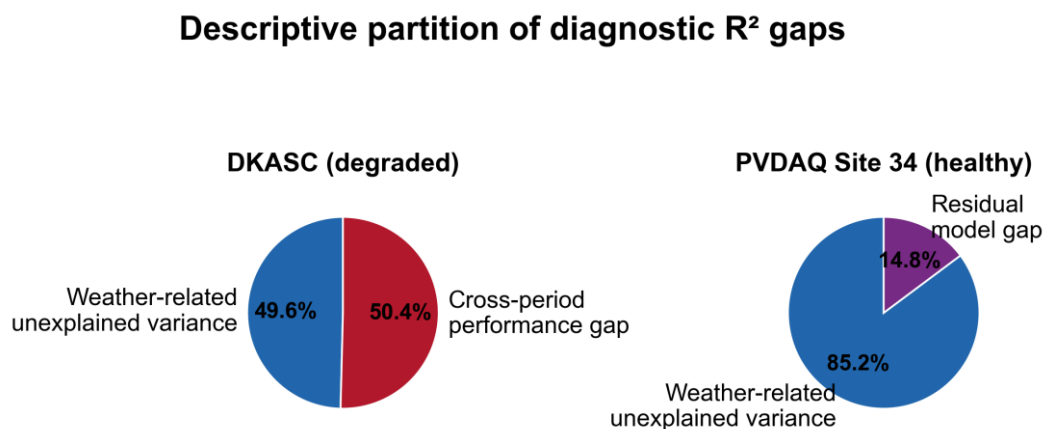


Fig. 5. Descriptive partition of selected diagnostic R^2 gaps. At DKASC, the within-period unexplained component and the cross-period performance gap contribute similar shares of the selected diagnostic total. At PVDAQ Site 34, the within-period unexplained component dominates and the remaining share is a residual model gap. Because $1 - R^2_{os}$ can contain weather variability, model misspecification, measurement noise, and omitted system-state information, these shares are descriptive summaries rather than causal variance components.

Diagnostic component	Formula	DKASC Value	DKASC Share (%)	PVDAQ Value	PVDAQ Share (%)
Weather-related unexplained variance	$1 - R^2_{os}$	0.621	49.6%	0.145	~85%
Cross-period performance gap	$R^2_{os} - R^2_{oc}$	0.630	50.4%	~0	~0%
Residual model gap	$R^2_{oc} - R^2_{model}$	~0	0%	small	~15%

Table 6. Descriptive partition of selected diagnostic R^2 gaps. The component $1 - R^2_{os}$ is a within-period unexplained term and is not uniquely attributable to weather; the quantities should not be interpreted as a causal error decomposition.

5.5 Configuration Sensitivity and Supporting Strategy Results

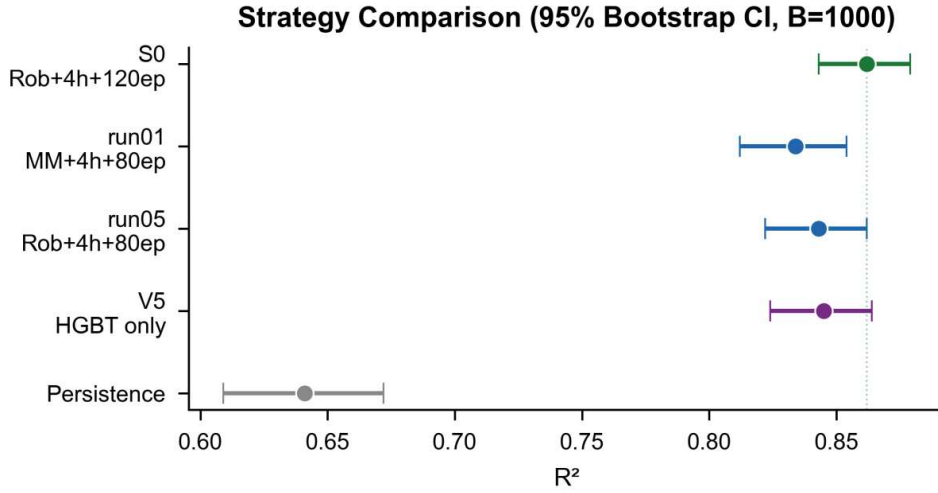


Fig. 6. Configuration sensitivity with descriptive 95% bootstrap CI (B=1,000) and nominal Diebold–Mariano comparisons. S0 achieves $R^2 = 0.862$ [0.843, 0.879]. ★ denotes nominal DM $p < 0.05$ versus S0.

Strategy	R ²	95% CI	RMSE (kW)	MAPE (%)	DM p	Sig.
S0: Rob+4h+120ep (Best)	0.862	[0.843, 0.879]	0.527	18.1	—	—
run01: MM+4h+80ep	0.834	[0.812, 0.854]	0.590	19.3	<0.001	★
run05: Rob+4h+80ep	0.843	[0.822, 0.862]	0.565	18.9	0.029	★
V5: HGBT primary	0.845	[0.824, 0.864]	0.558	17.9	—	—
Persistence baseline	0.641	[0.609, 0.672]	0.851	25.3	—	—
XGBoost [46]	0.82–0.85	—	—	—	—	Lit.
LightGBM [46]	0.83–0.86	—	—	—	—	Lit.
TCN-GRU [49]	0.85–0.88	—	—	—	—	Lit.

Table 7. Configuration sensitivity and external benchmark context. Nominal DM p values are reported only where configurations were directly evaluated against S0; literature ranges are contextual and are not controlled comparisons.

External benchmark context. Table 7 also lists R^2 ranges reported for XGBoost [46], LightGBM [46], and TCN-GRU [49] in other PV forecasting studies. These serve as context only; differences in datasets, horizons, and preprocessing preclude a controlled performance comparison.

Scope of the strategy analysis. Table 3 defines six candidate adaptation strategies, whereas Table 7 reports only the configurations for which directly comparable metrics and confidence intervals were available. Results for S1–S5 are therefore supporting analyses rather than a complete controlled ranking. S4 underlies the retrospective capacity-feature diagnostic of Section 5.2, and S5 produced $\Delta R^2 = -0.050$ at DKASC (DM $p < 0.01$).

5.6 Stratified Error Analysis

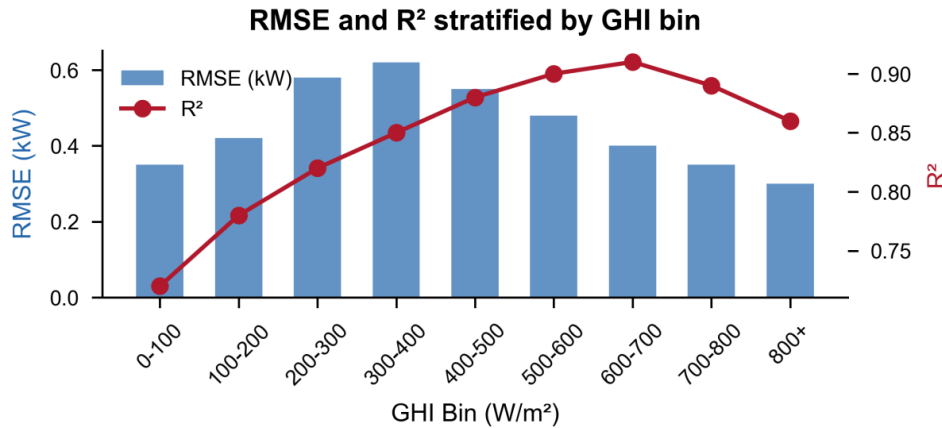


Fig. 7. RMSE and R^2 stratified by GHI bin. Peak RMSE in the 200–500 W/m² transition zone is consistent with combined cloud variability and capacity shift, but the stratified pattern is descriptive and does not isolate either mechanism.

5.7 Stable-Site Contrast: PVDAQ Site 34 (14-Year Record)

To examine behaviour at a comparatively stable site, we applied the Oracle-Persistence procedure to the 14-year PVDAQ Site 34 record (2010–2023); the site was chosen for climatic contrast before its long-term trend was quantified.

Over the 14-year record, the effective conversion efficiency η_{eff} showed no statistically significant linear trend ($-0.13\% \text{ yr}^{-1}$, $p = 0.35$; 95% confidence interval spanning zero). The Pettitt test likewise identified no structural break ($p = 0.41$), and P95 power remained within 10.1–11.0 kW. These results are consistent with a comparatively stable system over the analysed period.

At PVDAQ Site 34, Oracle-Cross $R^2 = 0.903$ exceeded $R^2_{\text{pers}} = 0.590$ and remained positive, so neither transfer-failure warning was triggered, and the KS statistic was only $D = 0.045$. Together with the non-significant long-term trend and Pettitt test, these results indicate stable effective capacity over the analysed period.

PVDAQ Site 34 exhibited a 0.60 kW range in monthly mean residuals, including a tendency to under-predict in August. This seasonal structure may reflect weather regimes, temperature effects, or unmodelled operational factors; it should not be read as confirmed recurring capacity drift without additional system-state evidence.

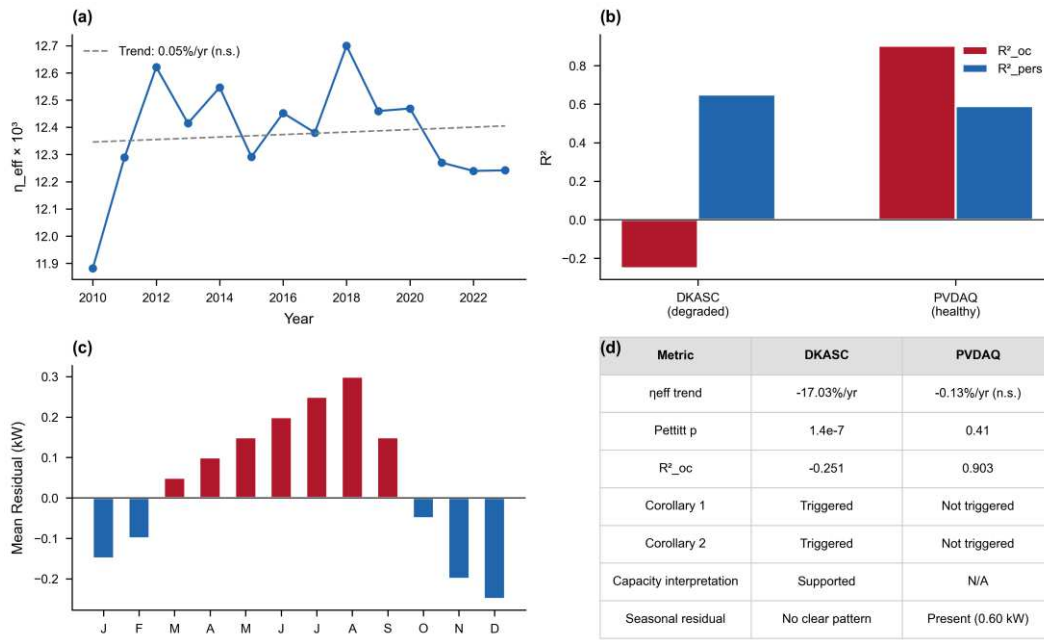


Fig. 8. Stable-site contrast at PVDAQ Site 34: (a) 14-year effective-efficiency trend ($-0.13\% \text{ yr}^{-1}$, not significant); (b) Oracle diagnostic comparison with DKASC; (c) seasonal residual pattern (0.60 kW range); and (d) dual-site diagnostic summary.

5.8 Future-Irradiance Sensitivity Analysis: Equation (2)

The 14-year PVDAQ record yields a 1-hour POA autocorrelation of $r = 0.780$ and $\text{Var}(\Delta\text{POA}_{1\text{h}}) \approx 53,644 \text{ (W/m}^2\text{)}^2$ — roughly $24.5\times$ the DKASC estimate ($\sim 800 \text{ (W/m}^2\text{)}^2$). Equation (2) then predicts that NWP benefit should be neutral-to-positive at PVDAQ and near-zero or harmful at DKASC.

At DKASC, integrating the evaluated weather-forecast features reduced accuracy ($\Delta R^2 = -0.050$, DM $p < 0.01$) — a real-data result consistent with Equation (2)'s prediction that additional weather information has little marginal value where short-horizon irradiance is strongly persistent. For PVDAQ Site 34, co-located operational NWP was unavailable, so we assessed sensitivity to imperfect future irradiance using calibrated synthetic noise; this directional sensitivity analysis ($\Delta R^2 = 0$ to $+0.004$) is reported in full in Supplementary Material S3. The opposite signs at the two sites are directionally consistent with the climate-dependence hypothesis, but the synthetic PVDAQ input precludes a direct operational-NWP validation claim.

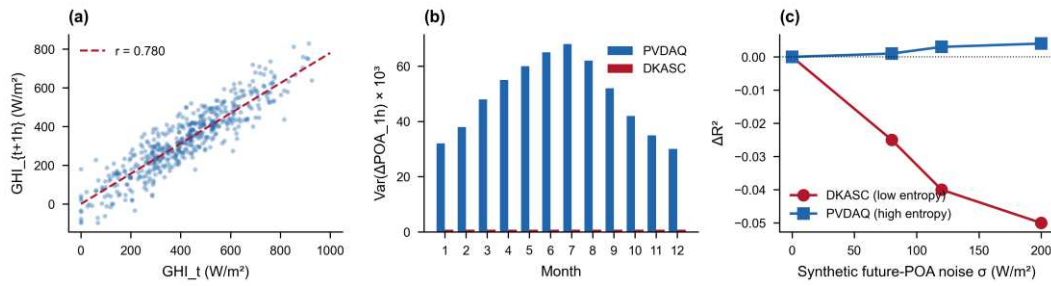


Fig. S3 (Supplementary Material). Future-irradiance sensitivity analysis: (a) 1-h POA autocorrelation at PVDAQ; (b) monthly $\text{Var}(\Delta\text{POA})$; and (c) change in R^2 across synthetic noise levels. The PVDAQ experiment uses synthetic future-irradiance information rather than operational NWP.

6. Discussion

6.1 Interpreting the Diagnostic R^2 Gaps

The selected diagnostic R^2 gaps were of comparable magnitude at DKASC, whereas the within-period unexplained term dominated at PVDAQ Site 34. This contrast is consistent with effective-capacity information becoming more informative when historical weather-to-power relationships fail to transfer. However, $1 - R^2_{\text{os}}$ also contains model misspecification, measurement noise, and omitted variables; the partition is descriptive and depends on the chosen models, features, and evaluation periods.

6.2 Evidence from the Two-Site Contrast

The dual-site design offers an initial contrast between pronounced cross-period transfer failure at DKASC and stable transfer at PVDAQ Site 34. Because only one strongly degraded and one stable site were analysed, these results demonstrate feasibility rather than population-level sensitivity or specificity.

6.3 Climate Dependence of Future-Weather Information

The synthetic future-irradiance experiment is consistent with the hypothesis that the marginal value of weather information depends on short-horizon irradiance variability. It does not, however, establish that NWP is generally harmful at desert sites or beneficial at more variable ones. Operational forecast archives are needed before this hypothesis can guide investment decisions.

The synthetic experiment provides directional sensitivity evidence only. Its absolute magnitudes and operational relevance require validation with real forecast products that preserve lead-time-dependent bias, spatial mismatch, and correlated errors.

6.4 Simplicity over Complexity

Among the directly comparable configurations in Table 7, S0 achieved the highest R^2 . Replacing MinMaxScaler with RobustScaler produced the largest statistically supported improvement ($\text{DM} =$

-4.73 , nominal $p < 0.001$). This indicates that robust preprocessing can be consequential under capacity-shifted power distributions, although it does not establish superiority across sites or architectures. More broadly, short-term forecasting models may require monitoring for changes in the weather-to-power relationship during multi-year deployment; the Oracle-Persistence comparison offers one candidate diagnostic, although prospective validation and strictly pre-origin capacity estimates are required before operational use.

7. Limitations and Future Work

This study has five principal limitations that define the agenda for future research:

Limitation 1 — Single degraded site. The degradation analysis relies on a single site (DKASC), so the 50/50 error split and the -17% /year degradation rate are site-specific. Multi-site validation across diverse climates and degradation mechanisms is needed.

Limitation 2 — Synthetic future-irradiance information. Equation (2) was assessed by adding Gaussian noise to observed future irradiance rather than by using archived operational forecasts (this synthetic PVDAQ analysis is now reported in Supplementary Material S3). The result therefore characterises directional sensitivity, not operational NWP value.

Limitation 3 — Deterministic forecasts only. All strategies produce point forecasts. Probabilistic forecasting with drift-adaptive prediction intervals is a natural extension, and the Oracle-Persistence framework can be adapted to track interval coverage probability.

Limitation 4 — Proxy validity and responsiveness. The rolling P95 estimator for $\hat{C}(t)$ may lag discrete events by up to 30 days and can respond to cloudiness, seasonal irradiance, clipping, curtailment, and missing-data patterns as well as to physical capacity. Stable GHI and the cross-site contrast reduce, but do not eliminate, this confounding. The strictly pre-origin estimator removes target-interval leakage at some cost in responsiveness; inverter telemetry, I–V curve measurements, or irradiance-normalised clear-sky capacity estimates would provide stronger system-state evidence. The present result therefore supports a capacity-consistent interpretation rather than identification of a specific fault mechanism.

Limitation 5 — Temporal dependence and inferential calibration. The five-minute and monthly series are serially dependent. Conventional bootstrap intervals and nominal DM, Pettitt, and KS p values may therefore be too narrow or anti-conservative unless dependence is explicitly modelled. The reported

inferential quantities are interpreted descriptively; confirmatory evaluation should use blocked resampling, dependence-robust forecast comparison, and autocorrelation-aware change-point calibration.

Future-work priorities are to: (1) validate across 5+ degraded sites spanning diverse climates and failure mechanisms; (2) integrate real archived NWP products to quantify Equation (2); (3) compare rolling P95 with telemetry-based and irradiance-normalised capacity estimators; (4) repeat uncertainty and forecast-comparison analyses with dependence-aware methods; (5) extend the framework to probabilistic forecasting with drift-adaptive prediction intervals; and (6) package the diagnostic into a prospectively evaluated toolkit for operational PV-fleet monitoring.

8. Conclusions

This study examined PV power forecasting under time-varying effective capacity and introduced the Oracle-Persistence diagnostic, using long-term records from DKASC and PVDAQ Site 34. Five conclusions follow within the limits of the two-site retrospective design:

C1. DKASC exhibits severe compound degradation ($-17.03\%/yr$; structural break in March 2023, $p = 1.4 \times 10^{-7}$), whereas PVDAQ Site 34 is stable over 14 years ($-0.13\%/yr$, $p = 0.35$, n.s.) — providing a healthy-system contrast across 26 combined site-years of field data.

C2. At DKASC, Oracle-Cross $R^2 = -0.251$ indicated severe cross-period transfer failure. Adding a retrospective capacity feature raised it to 0.955 (a diagnostic upper bound), and a strictly pre-origin, leakage-free capacity estimate retained $R^2 = 0.873$ (well above persistence, 0.646). This supports a capacity-consistent interpretation under the stated information boundary, but it does not identify a physical mechanism or constitute prospective deployment validation.

C3. At PVDAQ Site 34, Oracle-Cross $R^2 = 0.903$ exceeded persistence and the long-term capacity indicators remained stable, providing a useful stable-site contrast rather than a population-level estimate of specificity.

C4. The best directly evaluated configuration achieved $R^2 = 0.862$ [0.843, 0.879]. Robust scaling produced the largest statistically supported improvement over MinMax scaling (Diebold–Mariano test), whereas the evaluated weather-forecast integration reduced performance at DKASC.

C5. The marginal value of future weather information appears climate-dependent: at DKASC, integrating the evaluated weather-forecast features reduced accuracy ($\Delta R^2 = -0.050$, DM $p < 0.01$), whereas a synthetic future-irradiance sensitivity analysis at PVDAQ Site 34 (Supplementary Material S3) was directionally positive. Real, archived NWP forecasts are required to test operational validity.

Overall, changes in effective system capacity can undermine the transfer of historical PV forecasting models. Comparing cross-period performance, a model-dependent within-period reference, and persistence provides a transparent diagnostic when the scores are computed under matched protocols and strictly pre-origin information boundaries. In this two-site study, the results support effective-capacity change as an important contributor to transfer failure at DKASC, while the PVDAQ contrast remains consistent with relative stability. Broader generalisability, causal attribution, and operational utility require multi-site prospective validation. To support adoption and reproducibility, the diagnostic code and processed data are available from the corresponding author upon reasonable request.

Ethics, Consent to Participate, and Consent to Publish

Ethics, Consent to Participate, and Consent to Publish declarations: not applicable.

Disclosure of Interest

No potential competing interest was reported by the authors.

Funding

No funding was received.

Data and Code Availability

The DKASC data used in this study are publicly available at <https://dkasolarcentre.com.au/>. The PVDAQ Site 34 data are available through the NREL Photovoltaic Data Acquisition (PVDAQ) Public Datasets record hosted by the Open Energy Data Initiative at <https://data.openei.org/submissions/4568> (DOI: <https://doi.org/10.25984/1846021>). The Oracle-Persistence diagnostic code, preprocessing scripts, and model configurations are available from the corresponding author upon reasonable request.

References

- [1] Gama, J. et al. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4), 1–37.
- [2] Jordan, D.C., Kurtz, S.R. (2013). Photovoltaic degradation rates—an analytical review. *Progress in Photovoltaics*, 21(1), 12–29.
- [3] DKA Solar Centre (2024). Desert Knowledge Australia Solar Centre. <https://dkasolarcentre.com.au/>
- [4] Deline, C. et al. (2021). Photovoltaic Data Acquisition (PVDAQ) Public Datasets. OEDI. DOI: 10.25984/1846021.
- [5] Pettitt, A.N. (1979). A non-parametric approach to the change-point problem. *J. Royal Stat. Soc. C*, 28(2), 126–135.
- [6] Diebold, F.X., Mariano, R.S. (1995). Comparing predictive accuracy. *J. Business & Economic Statistics*, 13(3), 253–263.
- [7] Yang, D. et al. (2018). History and trends in solar irradiance and PV power forecasting. *Solar Energy*, 168, 60–90.
- [8] Lorenz, E. et al. (2009). Irradiance forecasting for PV systems. *IEEE JSTARS*, 2(1), 2–10.
- [9] Zippenfenig, P. (2023). Open-Meteo.com Weather API. Zenodo. DOI: 10.5281/ZENODO.7970649.
- [10] IRENA (2024). Renewable Capacity Statistics 2024. International Renewable Energy Agency.
- [11] Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- [12] Vaswani, A. et al. (2017). Attention is all you need. *NeurIPS*, 30.
- [13] Pedregosa, F. et al. (2011). Scikit-learn: Machine learning in Python. *JMLR*, 12, 2825–2830.
- [14] Bifet, A., Gavaldà, R. (2007). Learning from time-changing data with adaptive windowing. *SIAM SDM*, 443–448.
- [15] Efron, B., Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall/CRC.
- [16] Lindig, S. et al. (2018). Review of statistical and analytical degradation models for photovoltaic modules and systems. *IEEE Journal of Photovoltaics*, 8(6), 1773–1786.
- [17] Phinikarides, A. et al. (2014). Review of photovoltaic degradation rate methodologies. *Renewable and Sustainable Energy Reviews*, 40, 143–152.
- [18] Skoczek, A. et al. (2009). The results of performance measurements of field-aged crystalline silicon photovoltaic modules. *Progress in Photovoltaics*, 17(4), 227–240.
- [19] Antonanzas, J. et al. (2016). Review of photovoltaic power forecasting. *Solar Energy*, 136, 78–111.
- [20] Voyant, C. et al. (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, 105, 569–582.
- [21] Das, U.K. et al. (2018). Forecasting of photovoltaic power generation and model optimization: A review. *Renewable and Sustainable Energy Reviews*, 81, 912–928.
- [22] Lu, J. et al. (2018). Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering*, 31(12), 2346–2363.
- [23] Webb, G.I. et al. (2016). Characterizing concept drift. *Data Mining and Knowledge Discovery*, 30(4), 964–994.
- [24] Pan, S.J., Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359.

- [25] Inman, R.H. et al. (2013). Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*, 39(6), 535–576.
- [26] Li, P. et al. (2020). A hybrid deep learning model for short-term PV power forecasting. *Applied Energy*, 259, 114216.
- [27] Zhou, H. et al. (2021). Short-term photovoltaic power forecasting based on long short-term memory neural network and attention mechanism. *IEEE Access*, 9, 55618–55627.
- [28] Qin, Y. et al. (2017). A dual-stage attention-based recurrent neural network for time series prediction. *Proceedings of IJCAI-2017*, 2627–2633.
- [29] Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of KDD 2016*, 785–794.
- [30] Truong, C. et al. (2020). Selective review of offline change point detection methods. *Signal Processing*, 167, 107299.
- [31] Deceglie, M.G. et al. (2018). RdTools: An open-source Python library for analyzing degradation in photovoltaic systems. *Proc. 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion*, 1–5.
- [32] Haque, A. et al. (2019). Fault diagnosis of photovoltaic modules. *Energy Science & Engineering*, 7(4), 1240–1255.
- [33] Köntges, M. et al. (2014). Review of Failures of Photovoltaic Modules. IEA PVPS Task 13, Report T13-01:2014.
- [34] Mathiesen, P., Kleissl, J. (2011). Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. *Solar Energy*, 85(5), 967–977.
- [35] Nespoli, A. et al. (2019). Day-ahead photovoltaic forecasting: A comparison of the most effective techniques. *Energies*, 12(9), 1621.
- [36] Srivastava, N. et al. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56), 1929–1958.
- [37] Kingma, D.P., Ba, J. (2015). Adam: A method for stochastic optimization. *Proceedings of ICLR 2015*.
- [38] Ruder, S. (2016). An overview of gradient descent optimisation algorithms. *arXiv:1609.04747*.
- [39] Wan, C. et al. (2015). Photovoltaic and solar power forecasting for smart grid energy management. *CSEE Journal of Power and Energy Systems*, 1(4), 38–46.
- [40] Akhter, M.N. et al. (2019). Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. *IET Renewable Power Generation*, 13(7), 1009–1023.
- [41] Fatemi, S.A. et al. (2016). Parametric methods for probabilistic forecasting of solar irradiance. *Renewable Energy*, 83, 836–848.
- [42] Wolff, B. et al. (2016). Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data. *Solar Energy*, 135, 197–208.
- [43] Shi, J. et al. (2012). Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *IEEE Transactions on Industry Applications*, 48(3), 1064–1069.
- [44] Theodorides, S. et al. (2020). Day-ahead hourly forecasting of power generation from photovoltaic installations using artificial neural networks. *Energies*, 13(18), 4864.
- [45] Engerer, N.A., Mills, F.P. (2014). KPv: A clear-sky index for photovoltaics. *Solar Energy*, 105, 679–693.
- [46] Gaboitaolelwe, J., Zungeru, A.M., Yahya, A. (2023). Machine learning based solar photovoltaic

power forecasting: A review and comparison. *IEEE Access*, 11, 56984–57006. DOI: 10.1109/ACCESS.2023.3270041

[47] Suárez-Cetrulo, A.L., Quintana, D., Cervantes, A. (2022). A survey on machine learning for recurring concept drifting data streams. *Expert Systems with Applications*, 213, 118934. DOI: 10.1016/j.eswa.2022.118934

[48] Xiang, Q., Zi, L., Cong, X. (2023). Concept drift adaptation methods under the deep learning framework: A literature review. *Applied Sciences*, 13(11), 6515. DOI: 10.3390/app13116515

[49] Li, Y., Song, L., Zhang, S. (2023). A TCN-based hybrid forecasting framework for hours-ahead utility-scale PV forecasting. *IEEE Transactions on Smart Grid*, 14(4), 2915–2928. DOI: 10.1109/TSG.2023.3236992

[50] Guermoui, M., Fezzani, A., Mohamed, Z. (2024). An analysis of case studies for advancing photovoltaic power forecasting through multi-scale fusion. *Scientific Reports*, 14, 57398. DOI: 10.1038/s41598-024-57398-z

[51] Yu, J., Li, X., Yang, L. (2024). Deep learning models for PV power forecasting: Review. *Energies*, 17(16), 3973. DOI: 10.3390/en17163973

[52] Hassan, M.A., Bailek, N., Bouchouicha, K. (2022). Evaluation of energy extraction of PV systems affected by environmental factors under real outdoor conditions. *Theoretical and Applied Climatology*, 150, 1155–1168. DOI: 10.1007/s00704-022-04166-6

[53] Gao, Y., Yao, J., Zhu, H. (2026). A novel CNN-LSTM photovoltaic power forecasting method integrating GMM-based error distribution modeling and correction. *Journal of Renewable and Sustainable Energy*, 18, 023501. DOI: 10.1063/5.0313815.

[54] Wang, M., Ma, X., Wang, R., Kari, T., Tang, Z. (2024). Short-term photovoltaic power prediction model based on hierarchical clustering of K-means++ algorithm and deep learning hybrid model. *Journal of Renewable and Sustainable Energy*, 16, 026102. DOI: 10.1063/5.0198444.

[55] Chen, G., Zheng, L. (2025). The short-term prediction method of PV output sequence based on the phase space reconstruction and GAT-LSTM hybrid model. *Journal of Renewable and Sustainable Energy*, 17, 063501. DOI: 10.1063/5.0281896.

[56] Li, B., Wang, H., Zhang, J. (2025). Short-term power forecasting of photovoltaic generation based on CFOA-CNN-BiLSTM-Attention. *Electrical Engineering*, 107(8), 10335–10347. DOI: 10.1007/s00202-025-03031-9.

[57] Hu, J. (2024). A short-term forecasting method for photovoltaic power generation based on the TCN-ECANet-GRU hybrid model. *Scientific Reports*, 14, 6744. DOI: 10.1038/s41598-024-56751-6.

[58] Zhao, L., Shen, Y. (2025). Proactive model adaptation against concept drift for online time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '25)*. DOI: 10.1145/3690624.3709210.

[59] Deng, Y., Liang, R., Fan, J., et al. (2025). Concept drift-aware time-series generation for online building load forecasting: An automated data augmentation paradigm. *ACM Transactions on Sensor Networks*. DOI: 10.1145/3737644.

[60] Azar, J., et al. (2025). Deep learning-based approach to automated monitoring of defects and soiling on solar panels. *Future Internet*, 17(10), 433. DOI: 10.3390/fi17100433.

[61] Jordan, D.C., Deceglie, M.G., et al. (2022). Photovoltaic fleet degradation insights. *Progress in Photovoltaics: Research and Applications*, 30(10). DOI: 10.1002/pip.3566.

[62] IEA PVPS (2025). PV Failure Fact Sheets: Degradation and Failure. IEA-PVPS Task 13, Report T13-30:2025.