International Journal of Advanced AI Applications
Online ISSN: 3104-9338
Print ISSN: 3104-932X

Quality-Aware Photovoltaic Forecasting under Missing Weather Data

Haitian Ren*

School of Energy, Belarusian National Technical University, Minsk, Belarus

Received: September 19, 2026

Revised: September 20, 2026

Accepted: September 22, 2026

Published online: September 22, 2026

To appear in: *International Journal of Advanced AI Applications,* Vol. 2, No. 10 (October 2026)

Corresponding Author:
Haitian Ren
(haitian.ren.ee@gmail.com)

Abstract. Operational photovoltaic forecasts often depend on irradiance and temperature measurements that may be interrupted by sensor or communication faults. This study develops Quality-Aware Dual-Path Conformal Photovoltaic Forecasting (QDPCC-PV), which combines a weather-aware Extra Trees expert with a power-only fallback expert. An availability gate changes their weights according to observed weather inputs, while physical clipping enforces feasible output and condition-specific conformal calibration provides prediction intervals. The method was evaluated chronologically on two measured photovoltaic plants sampled every 15 min at 15-, 60-, and 240-min horizons. Under complete inputs, mean absolute errors were 0.0376, 0.0463, and 0.0684 per unit, corresponding to improvements of 32.8%, 45.7%, and 65.3% over persistence. With 60% random weather loss at the 60-min horizon, QDPCC-PV obtained 0.0492 error, compared with 0.0664 for median imputation. The nominal 90% interval reached 91.8% coverage at 240 min but undercovered at shorter horizons because calibration and test residuals shifted over time. The results show that explicit input-quality routing protects point forecasts from weather-data loss, while adaptive calibration remains necessary for reliable short-horizon intervals.

Keywords: *sensor outages; extra trees; mixture of experts; interval calibration; multi-horizon prediction*

1. Introduction

The rapid growth of photovoltaic (PV) generation has made short-horizon forecasting essential for reserve scheduling, market bidding, battery dispatch, and voltage control. Yet operational forecasts often depend on irradiance and temperature streams that can fail because of sensor faults, maintenance, communication loss, or delayed weather services. In such cases, conventional imputation makes the input numerically complete without making it reliable, and a multivariate model may perform worse than a target-only model [1,2,15].

This study focuses on 15-, 60-, and 240-min horizons, spanning immediate control, intra-hour scheduling, and short-term planning. Existing physical, statistical, machine-learning, and deep temporal models improve clean-data accuracy in different settings [3-14], but most assume that exogenous inputs remain available at inference time. They also commonly report point forecasts alone, although operators need uncertainty estimates that respond to data quality.

These gaps motivate Quality-Aware Dual-Path Conformal Photovoltaic Power Forecasting (QDPCC-PV). The method uses a weather-aware expert when meteorological inputs are observed, progressively transfers weight to a power-only expert as those inputs disappear, enforces physical output bounds, and calibrates uncertainty separately by horizon and missingness condition. The resulting design isolates a practical question: can explicit routing provide graceful degradation without hiding the effect behind a high-capacity predictor?

The contributions of this work are summarized as follows.

• A dual-path forecasting structure is developed with a meteorology-aware expert and a power-only fallback expert. Unlike unconditional imputation, the structure remains explicitly aware of which variables are observed.

• A transparent quality gate converts the observed-feature ratio into expert weights. The mechanism is directly auditable and can later be replaced by a learned gate using gap duration, information age, and regime descriptors.

• Physical projection and missingness-conditioned split conformal prediction are integrated with the point forecast to produce feasible outputs and empirical uncertainty intervals.

• A reproducible chronological experiment evaluates complete data, random loss, block outages, and total weather loss using two measured PV plants and three horizons. Both positive and negative findings, including short-horizon interval undercoverage, are reported.

2. Related Work

2.1. Statistical and Machine-Learning Methods

Persistence assumes that future power equals the most recent observation. Although simple, it is a strong very-short-term baseline because PV output is highly autocorrelated. Linear and autoregressive models add trend and seasonal components but remain limited when weather-power relations change across irradiance and temperature regimes. Support vector regression maps the inputs into a nonlinear feature space, while decision-tree ensembles approximate nonlinear interactions without extensive scaling. Random Forest (RF) averages randomized decision trees to reduce variance [20]. Extra Trees (ET), formally Extremely Randomized Trees, additionally randomizes split thresholds and can reduce training cost and variance [21]. These properties make ET suitable for a transparent pilot implementation of a dual-expert architecture.

Tree ensembles are also useful when the available dataset is smaller than the training corpora normally used for deep temporal networks. They accept mixed nonlinear predictors, do not require gradient-based optimization, and expose feature-importance summaries that support diagnostic analysis. Their principal limitation is that temporal structure must be supplied explicitly through lags, rolling statistics, and calendar variables. They also extrapolate poorly outside the target range represented in training. The present study addresses the second limitation with a physical projection and accepts the first as a deliberate trade-off: the aim is to isolate the value of quality-aware routing before introducing a higher-capacity backbone. This separation is important because a complicated predictor can otherwise obscure whether robustness originates from the model architecture, the imputation rule, or the routing mechanism.

2.2. Deep Temporal Models

Deep temporal models such as LSTM, TCN, and Transformer architectures can learn long-range and multi-scale dependencies [5-10]. PVPNet is particularly relevant because it processes heterogeneous PV inputs through separate streams [1], while recent Transformer variants add physical features, frequency decomposition, or feature selection [11-13]. However, higher representational capacity does not by itself define how a forecaster should degrade when exogenous variables are missing. This study therefore uses the same Extra Trees family for both experts to isolate the value of the routing mechanism from backbone capacity and tuning effects; replacing the experts with deep models is left for future work.

2.3. Missing Data and Domain Shift

Numerical weather prediction (NWP) can differ systematically from local measured data. Domain-generalization and weather-correction methods attempt to learn representations that transfer across seasons or stations [22]. Missing-data research uses interpolation, matrix completion, generative imputation, mask-aware attention, transfer learning, and direct forecasting from incomplete covariates. Yan et al. [15] reported that covariates are useful only when their quality is sufficiently high. This result supports a conditional architecture in which the contribution of meteorological inputs depends on observed quality rather than remaining fixed.

Missingness must be distinguished by mechanism and duration. Independent random deletion approximates packet loss or isolated sensor errors, whereas block deletion represents communication outages, maintenance periods, or frozen acquisition systems. Missing not at random is more difficult because the probability of absence may depend on the weather regime or the sensor value itself; for example, a pyranometer may saturate or fail preferentially during extreme conditions. Simple interpolation is reasonable for very short smooth gaps but can suppress ramps across longer gaps. Statistical imputation expresses a central tendency but understates uncertainty, and learned imputation can transfer bias from the reconstruction model into the forecaster. The proposed design does not claim to reconstruct the missing weather state. Instead, it explicitly reduces reliance on the affected expert and retains a valid power-only route.

2.4. Probabilistic Forecasting

Quantile regression directly estimates conditional quantiles but requires a separate loss design and can produce crossing quantiles. Parametric methods specify distributions such as Gaussian or mixture densities; misspecification can yield unreliable tails. Large-scale energy forecasting competitions have reinforced the need to evaluate both accuracy and uncertainty [23]. CP instead ranks residual nonconformity scores on a held-out calibration set. Conformalized quantile regression combines learned quantiles with conformal correction [17], while dynamic time-series conformal methods relax some independence assumptions [18]. Renkema et al. [24] found that K-nearest-neighbor and Mondrian conformal variants improve photovoltaic interval quality. Adaptive conformal inference updates the error rate online under distribution shift [19]. QDPCC-PV uses a simpler missingness-conditioned split method so that its limitations can be measured directly. Table 2 positions QDPCC-PV against representative studies in terms of missing-weather handling and interval prediction.

The attraction of conformal prediction is model agnosticism: the point predictor can be changed without redesigning the interval construction. Its finite-sample marginal guarantee, however, depends on exchangeability between calibration and future nonconformity scores. Chronological solar data violate ideal exchangeability through seasonality, weather regimes, equipment changes, and autocorrelated ramps. A single global residual quantile can also mix daytime and nighttime cases or complete and incomplete inputs. Conditioning calibration on horizon and missingness reduces one source of heterogeneity, but it does not remove temporal drift. For this reason, the empirical coverage reported later is treated as an evaluated performance measure rather than assumed from theory. Undercoverage is retained and discussed because hiding it would give a misleading impression of operational reliability.

Table 1. Symbols and abbreviations.

Symbol	Definition	Unit
PV	Photovoltaic	-
ML	Machine learning	-
DL	Deep learning	-
ET	Extra Trees	-
CP	Conformal prediction	-
PICP	Prediction interval coverage probability	-
MPIW	Mean prediction interval width	p.u.
p_t	Normalized PV power at time t	p.u.
I_t	Solar irradiance	dataset unit
T_t^{amb}	Ambient temperature	°C
T_t^{mod}	Module temperature	°C
m_t,j	Availability indicator for feature j	0 or 1
g_t	Meteorological-expert gate weight	0-1
h	Forecast horizon	steps
L	Look-back length	steps
α	Target miscoverage rate	-

Table 2. Comparison with related studies.

Study	Core method	Weather loss	Intervals
Huang and Kuo [1]	Multi-input 1D CNN	No	No
Tao et al. [12]	Physics-aided Transformer	No	No
Yan et al. [15]	Missing-aware transfer	Yes	No
Renkema et al. [24]	Regression plus CP	No	Yes
El-Kenawy et al. [11]	PatchTST plus SMOA	No	No
QDPCC-PV	Quality-gated dual ET	Yes	Yes

3.1. Input and Output

Consider a plant-level sequence sampled every Δt = 15 min. The normalized active power at time t is p_t. The weather vector is $x_t = [I_t, T_t^{\text{amb}}, T_t^{\text{mod}}]$, where I_t is irradiance, T_t^{amb} is ambient temperature, and T_t^{mod} is module temperature. The binary vector m_t has dimension d = 3; m_t,j equals one when weather variable j is observed and zero when it is missing. The model uses a power history of L = 192 steps, corresponding to 48 h, and produces a point prediction and prediction interval for horizon h. Table 1 defines the principal symbols and abbreviations used throughout the formulation.

$$\hat{p}_{t+h}, C_{t+h} = F(p_{t-L+1:t}, x_t, m_t, h) \tag{1}$$

Here F denotes the complete forecasting operator, $p_{t-L+1:t}$ is the historical power window, C_{t+h} is the interval forecast, and h belongs to {1,4,16}, corresponding to 15, 60, and 240 min. The target power is normalized to per unit (p.u.) by a robust plant capacity proxy defined as the 99.9th percentile of observed plant AC power. This choice limits sensitivity to isolated maximum values while retaining a physically interpretable scale.

3.2. Missing-Data Conditions

Three input-quality states are considered. Complete data contain all weather channels. Missing completely at random removes each weather entry independently with probability ρ. Block missingness removes consecutive segments of all weather variables for 2-8 h, representing communication or sensor outages. Total loss sets ρ = 1 and forces the system to operate only from historical power and time variables. Artificial deletion is applied only to calibration and test covariates; the target sequence is never deleted.

3.3. Objectives

The point-forecast objective is to minimize expected absolute or squared error while maintaining stability as ρ increases. The probabilistic objective is to construct an interval whose empirical coverage approaches 1-α while remaining narrow. These objectives can conflict: increasing interval width improves coverage but reduces sharpness. Accordingly, both prediction interval coverage probability and mean prediction interval width are reported.

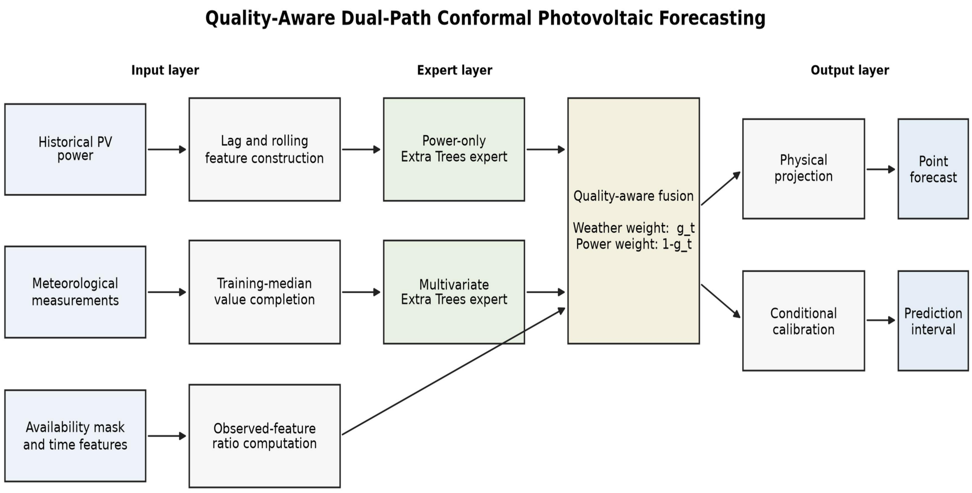

Figure 1. QDPCC architecture.

4. Proposed Method

4.1. Feature Construction

Figure 1 summarizes the dual-path architecture and its routing, projection, and calibration stages. The power-only feature vector s_t contains seven direct lags, rolling means and standard deviations, and periodic calendar encodings. Lags 1, 2, 4, 8, and 16 represent the previous 15 min to 4 h. Lags 96 and 192 represent the same time one and two days earlier. Rolling means and standard deviations are computed over 4, 16, and 96 steps, corresponding to 1 h, 4 h, and 24 h. Calendar variables include sine and cosine encodings of time of day and day of year. Cyclic encoding avoids an artificial discontinuity between 23:45 and 00:00.

$$s_t = [p_{t-1}, p_{t-2}, p_{t-4}, p_{t-8}, p_{t-16}, p_{t-96}, p_{t-192}, \boldsymbol{\mu}_t, \boldsymbol{\sigma}_t, c_t] \tag{2}$$

In (2), μ_t and σ_t are vectors of rolling means and standard deviations, and c_t is the cyclic calendar vector. The multivariate feature vector z_t augments s_t with the three weather variables.

$$z_t = \left[s_t, I_t, T_t^{\mathrm{amb}}, T_t^{\mathrm{mod}}\right] \tag{3}$$

Missing weather entries are filled with medians computed from the training split. This operation supplies a valid numeric input to the tree ensemble, but the mask is retained and affects the gate. Consequently, a filled value is never treated as an observed value by the complete system.

All temporal predictors are causal. A lag indexed by k contains power measured at t-k, and every rolling statistic is calculated after shifting the power series by one sample. Thus, the rolling window ending at forecast origin t never includes the target $p_{t+\mathrm{h}}$. Calendar variables are known in advance and do not create leakage. The same feature schema is built before the chronological split, but any statistic requiring fitted information, including weather medians, is estimated only from the training interval. This distinction prevents an apparently minor preprocessing step from transferring information from calibration or test periods into training. At inference, the feature builder also records whether each weather value was observed, imputed, or unavailable so that the routing decision can be audited independently of the numerical values supplied to the regressor.

4.2. Power-Only Expert

The power-only expert f_P is an ET regressor. Each tree is fitted to a bootstrap-free randomization of samples and features, and its split thresholds are selected randomly from candidate ranges. The ensemble prediction is the arithmetic mean of all tree outputs. This expert does not receive irradiance or temperature, so its input distribution is unchanged when all weather channels fail. It acts as a fail-operational fallback rather than as a competing final model.

4.3. Weather-Aware Expert

The multivariate expert f_M uses the same lag and time features as f_P and additionally receives irradiance, ambient temperature, and module temperature. Irradiance provides the strongest instantaneous physical relation to PV output, whereas temperature captures efficiency and thermal-state effects. Using the same ET family for both experts isolates the value of routing from differences in model capacity. Each expert contains 180 trees, uses a minimum leaf size of two, and considers 80% of the features at each split.

4.4. Quality Gate

The quality gate is the fraction of weather channels observed at the current forecast origin. With d weather variables, the gate is defined by (4). When all variables are present, g_t = 1 and the multivariate expert is used. When all are absent, g_t = 0 and the power-only expert is used. Partial availability produces a convex combination.

$$g_t = \frac{1}{d}\sum_{j=1}^{d} m_{t,j}, 0 \le g_t \le 1 \tag{4}$$

$$\hat{p}^F_{t+h} = g_t f_M(z_t) + (1 - g_t) f_P(s_t) \tag{5}$$

The deterministic gate is deliberately interpretable. It avoids training an additional network on a dataset that spans only one month. For larger datasets, g_t can be learned from the mask, consecutive gap length, information age, disagreement between measured and forecast irradiance, weather regime, and forecast horizon. The deterministic form therefore represents the minimum auditable implementation of the broader quality-aware concept.

The convex fusion has several useful boundary properties. If the three weather variables are observed, the output equals the weather-aware expert; if none is observed, the output equals the power-only expert; and for partial availability the forecast remains between the two expert outputs before physical clipping. Because the gate depends only on availability, it cannot improve the forecast by inspecting the unknown target. It also avoids abrupt switching when one of several channels disappears. Nevertheless, equal weighting of weather variables is a simplification. Irradiance normally contains more direct predictive information than ambient temperature, and a stale observation should not receive the same confidence as a current observation. A learned gate should therefore be constrained, calibrated, and tested under deliberately shifted conditions rather than optimized only for average clean-data accuracy.

4.5. Physical Projection

Tree ensembles generally remain within the target range observed during training, but fusion and future extensions may produce negative or above-capacity values. The physical projection in (6) clips the normalized forecast to [0,1]. This constraint represents nonnegative production and the plant-capacity limit. It does not model curtailment, cloud enhancement, inverter clipping, or detailed module physics.

$$\hat{p}_{t+h} = min\left(1, max\left(0, \hat{p}^{F}_{t+h}\right)\right) \tag{6}$$

4.6. Conditional Conformal Calibration

After both experts are trained, their parameters are frozen. The chronological calibration split is passed through the same missingness condition used at test time. For each horizon h and condition c, the absolute error is used as a nonconformity score.

$$s_{i,h,c} = |p_{i+h} - \hat{p}_{i+h}| \tag{7}$$

$$q_{h,c} = Q_{1-\alpha}\left(\{s_{i,h,c}: i \in \mathcal{I}_{\mathrm{cal},c}\}\right) \tag{8}$$

$Q_{1-\alpha}$ is the finite-sample empirical quantile and I_cal,c is the set of calibration indices under condition c. With α = 0.10, the nominal interval level is 90%. The lower and upper limits are physically clipped.

$$C_{t+h} = \left[max(0, \hat{p}_{t+h} - q_{h,c}), min(1, \hat{p}_{t+h} + q_{h,c})\right] \tag{9}$$

Conditioning on missingness allows the interval to widen when input quality decreases. It does not guarantee conditional coverage under arbitrary temporal drift. The empirical results quantify this distinction.

Calibration is performed after point-model training. For each plant, horizon, missingness pattern, and missing rate, the calibration covariates are corrupted by the same generator used to define the corresponding test condition. QDPCC-PV produces calibration forecasts, and their absolute residuals form the nonconformity sample. The finite-sample quantile uses the ceiling-adjusted rank associated with (1-α)(n+1), where n is the number of calibration residuals. At prediction time, the selected quantile is added to and subtracted from the point forecast, after which both limits are clipped to the physical range. No test residual is included in the quantile. This procedure makes the interval reproducible and prevents the common error of tuning interval width on the evaluation set.

4.7. Procedure

Training and inference consist of seven stages: aggregate inverter power; construct lags and rolling statistics; create chronological splits; train the two experts; generate missingness-conditioned calibration residuals; compute the quality-aware fused forecast; and evaluate point and interval metrics. No test observation is used for model fitting, feature scaling, median calculation, or interval quantile estimation. Figure 2 presents this end-to-end experimental workflow.

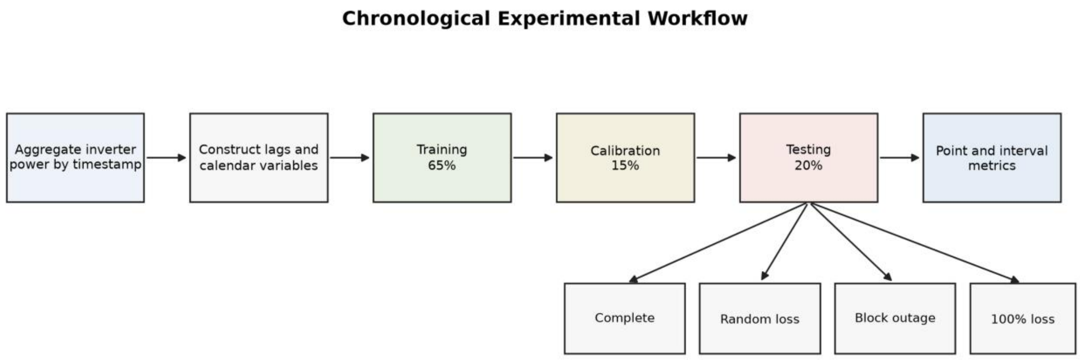

Figure 2. Experimental workflow.

5. Experiments

5.1. Datasets

Experiments use the public Solar Power Generation Data collection [25]. It contains two grid-connected PV plants in India, each with 22 inverter identifiers and a plant-level weather sensor. Generation and weather files are sampled every 15 min between May and June 2020. Available electrical variables include direct-current power, alternating-current power, daily yield, and total yield. Available meteorological variables include irradiance, ambient temperature, and module temperature. Figure 3 shows representative five-day power and weather profiles, and Table 3 summarizes the two plant records.

Generation data were aggregated to the plant level by summing inverter alternating-current power at each timestamp. Weather observations repeated across inverter rows were averaged. A complete 15-min timestamp index was constructed. Short weather gaps in the clean reference sequence were linearly interpolated for up to four observations and then forward/backward filled at the boundaries. Missing plant power timestamps were set to zero only when the aligned plant record contained no generation observation. After the 192-step history requirement, 3,072 samples remained for each plant, covering 17 May to 17 June 2020.

The public files do not provide a complete engineering description of module technology, tilt, azimuth, inverter clipping logic, curtailment commands, maintenance events, or sensor calibration. Consequently, the experiment is a data-driven plant-level study rather than a detailed physical simulation. The two plants also cover the same limited seasonal window, so cross-season and cross-climate generalization cannot be inferred. These constraints motivate the chronological design and the conservative interpretation of the results. The dataset is nevertheless suitable for the present question because it provides synchronized generation and meteorological measurements, multiple inverter streams, repeated daily cycles, and enough naturally variable conditions to construct controlled weather-input outages without altering the target observations. The feature-correlation structure used to support predictor selection is shown in Figure 4.

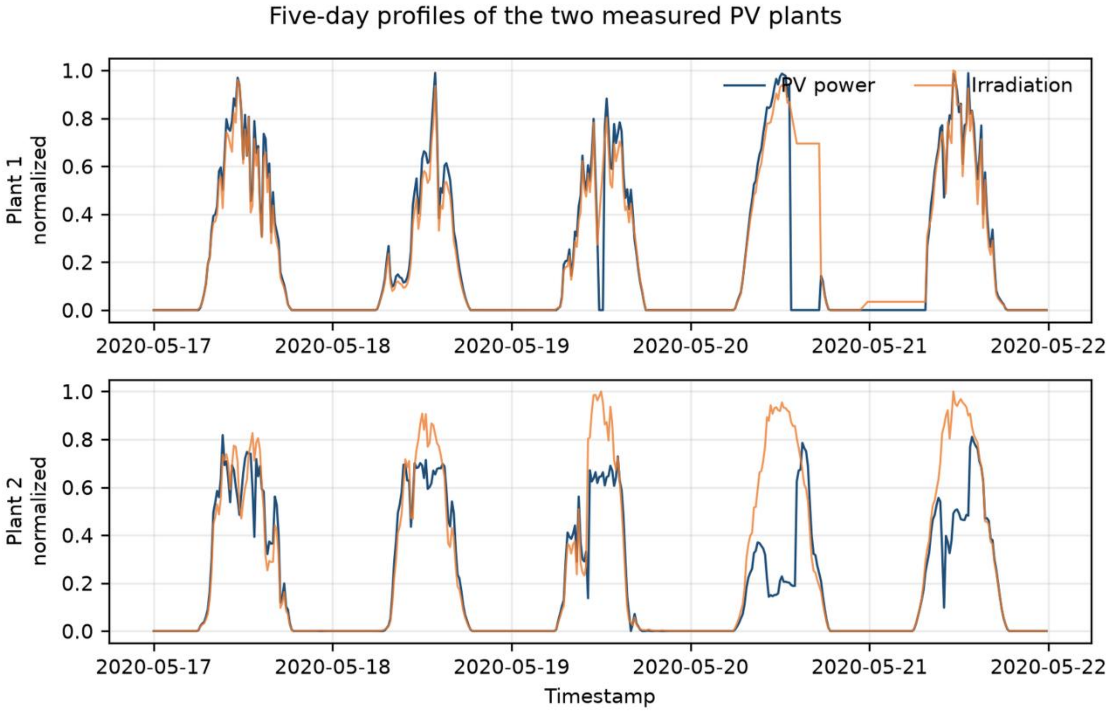

Figure 3. Five-day PV and weather profiles.

Table 3. Dataset summary.

Plant	Samples	Period	Capacity proxy	Daylight
1	3,072	May 17-Jun. 17	27,886.8 kW	34.6%
2	3,072	May 17-Jun. 17	24,440.5 kW	51.5%

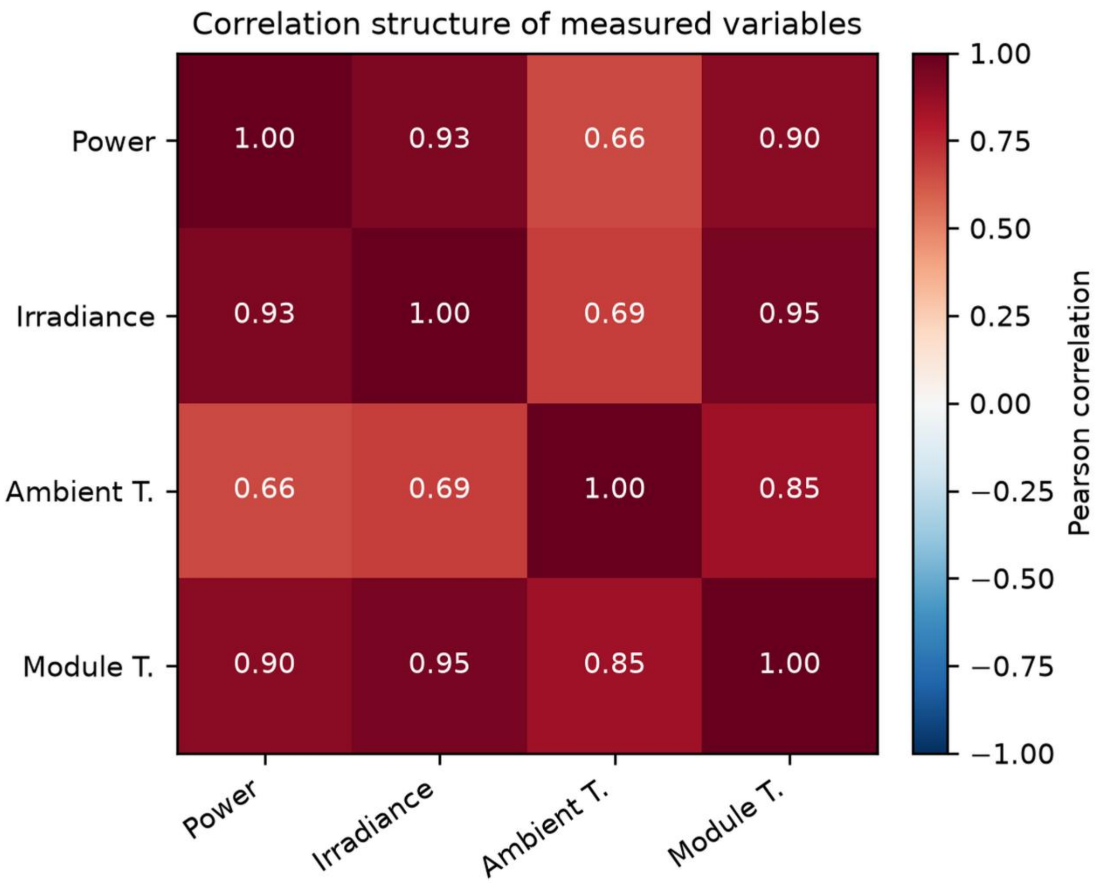

Figure 4. Feature correlation matrix.

5.2. Baselines

Four forecasting configurations are compared. Persistence sets the future prediction equal to the most recent power observation. Power-only ET uses the historical-power and calendar features described in Section 4.1. Median-impute ET uses all features but fills missing weather values with training medians and does not change model weight according to availability. QDPCC-PV combines the two experts using (5). Because the study focuses on missingness routing, identical ET settings are used for the two learned baselines and both QDPCC-PV experts.

PVPNet, LSTM, PatchTST, and the PatchTST-Swordfish model are discussed as literature references but are not assigned numerical results in this pilot because an honest comparison would require reimplementation, multiple random seeds, and the longer training periods used in their original studies. Fabricated or copied metrics from different datasets would not constitute a valid baseline.

5.3. Preprocessing

Timestamps were sorted before any split was performed. The first 65% of each plant sequence was used for training, the next 15% for conformal calibration, and the final 20% for testing. Sliding-window samples that required information outside their own chronological split were excluded. All medians and the capacity proxy were obtained without using future target values. Time-of-day and day-of-year features were encoded with sine and cosine functions. Power lags and rolling statistics used a one-step shift, preventing the current or future target from entering the feature vector.

Power was normalized separately for each plant by a high quantile of observed alternating-current output rather than by the absolute maximum. This reduces sensitivity to isolated spikes while retaining a comparable per-unit scale. Nighttime observations were retained because zero-output periods are part of operational forecasting, but performance is also interpreted in light of the resulting class imbalance between easy nighttime cases and more variable daylight cases. The corruption generator modifies only meteorological predictors; historical power lags, calendar variables, and targets remain unchanged. This isolates the effect of weather availability and prevents a mixed experiment in which missing targets, missing covariates, and target reconstruction are evaluated simultaneously.

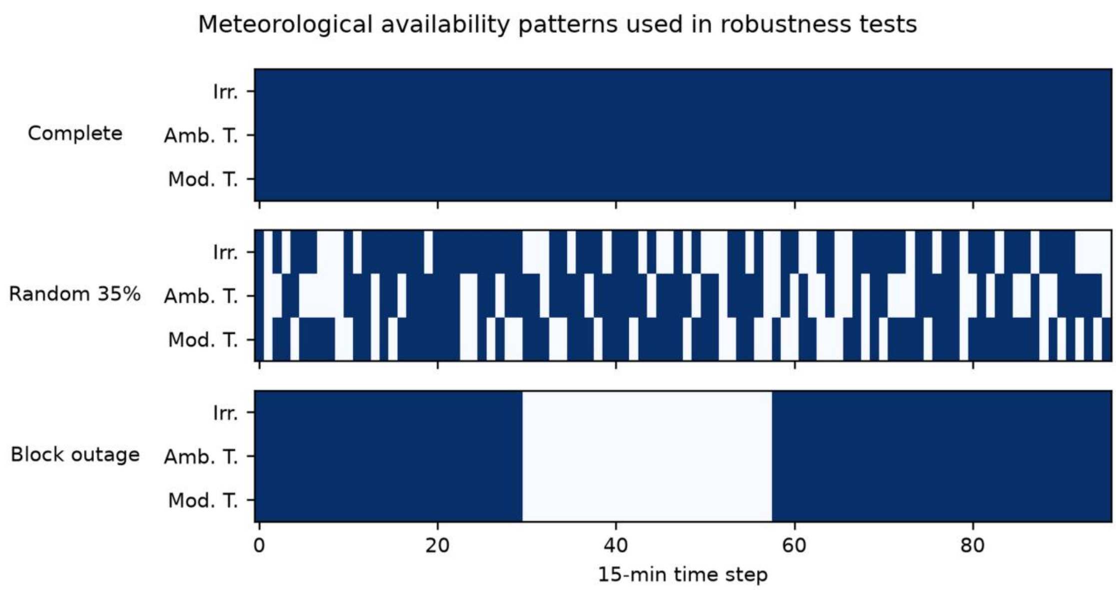

Figure 5. Simulated missing-data patterns.

Figure 5 illustrates the random, block, and total-loss patterns used in the corruption experiments. Random missingness rates were 0%, 30%, 60%, and 100%. For block experiments, contiguous segments between 8 and 32 samples, equivalent to 2-8 h, were repeatedly removed until the requested missing fraction was reached. A fixed random generator seed, 20260917, was used to make the corruption experiment reproducible.

5.4. Training

Each horizon and plant was trained independently. ET regressors contained 180 trees, minimum leaf size two, and a maximum-feature fraction of 0.8. Trees were trained in parallel using all available processor cores. The random state was deterministically derived from the plant identifier and forecast horizon. These settings were selected as a stable computational configuration rather than by optimizing against the test set. The point model was fitted only on clean training data. Calibration and testing were repeated for each missingness pattern and rate.

The independent horizon design avoids recursive error accumulation: a 240-min forecast is trained directly for that lead time rather than obtained by repeatedly applying a 15-min model. It also allows the relative value of weather and lag features to change with horizon. The training procedure does not use early stopping because ET builds a fixed number of randomized trees, and no validation score is used to choose among configurations. The calibration split has a single purpose: estimating residual quantiles after the point models are frozen. This separation is stricter than using one validation interval for both hyperparameter selection and uncertainty calibration. For larger benchmark experiments, a nested chronological validation period should be introduced before the calibration block so that hyperparameters, interval quantiles, and final test metrics remain statistically distinct.

Table 4. Experimental settings.

Parameter	Value	Meaning
Δt	15 min	Sampling interval
L	192	48-h look-back
h	1, 4, 16	15-, 60-, 240-min horizons
Trees	180	Trees per expert
Leaf size	2	Minimum samples per leaf
Feature fraction	0.8	Candidate-feature fraction
Train/cal/test	65/15/20%	Chronological partition
ρ	0, .3, .6, 1	Weather missing rate
α	0.10	Nominal miscoverage
Seed	20260917	Corruption reproducibility

5.5. Metrics

Mean absolute error (MAE) measures the average magnitude of point errors and is less sensitive to large ramps than squared error. Root mean squared error (RMSE) emphasizes large deviations. The coefficient of determination R^2 compares residual variance with the variance of observed power. Let N be the number of test observations, p_n the observed normalized power, $\hat{p}_n$ the point forecast, and p-bar the sample mean.

$$\text{MAE} = \frac{1}{N}\sum_{n=1}^{N} |p_n - \hat{p}_n| \tag{10}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} (p_n - \hat{p}_n)^2} \tag{11}$$

$$R^2 = 1 - \frac{\sum_{n=1}^{N} (p_n - \hat{p}_n)^2}{\sum_{n=1}^{N} (p_n - \bar{p})^2} \tag{12}$$

Prediction interval coverage probability (PICP) is the proportion of targets lying between lower limit l_n and upper limit u_n. Mean prediction interval width (MPIW) measures sharpness. High-quality intervals should attain coverage close to the nominal level without becoming unnecessarily wide.

$$\text{PICP} = \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}\,(l_n \le p_n \le u_n) \tag{13}$$

$$\text{MPIW} = \frac{1}{N}\sum_{n=1}^{N} (u_n - l_n) \tag{14}$$

Metrics are first calculated separately for each plant and then averaged, preventing the plant with more usable test samples from dominating the result. All power errors and widths are expressed in per unit. The present pilot uses one deterministic seed for model fitting; therefore, the reported table should not be interpreted as a multi-seed significance study.

MAE and RMSE answer different operational questions. MAE describes typical absolute deviation, whereas RMSE assigns greater influence to rare large errors that may drive reserve requirements. R^2 is included to show explained variation but can be misleading when the test segment contains many nighttime zeros. PICP and MPIW must be read together: a very wide interval can achieve high coverage without being useful, while a narrow interval with low coverage is overconfident. The reported metrics therefore support complementary conclusions rather than a single ranking. A production study should additionally report daylight-only errors, normalized ramp error, pinball loss or weighted interval score, inference latency, and confidence intervals across sites and seeds.

6. Results and Discussion

6.1. Clean-Input Accuracy

Table 5 reports performance with complete meteorological information. QDPCC-PV assigns $g_t = 1$ in this condition and consequently equals the multivariate expert. At 15 min, its MAE was 0.0376 p.u., compared with 0.0560 for persistence and 0.0426 for power-only ET. At 60 min, its MAE was 0.0463, and at 240 min it was 0.0684. Relative to persistence, the reductions were 32.8%, 45.7%, and 65.3%, respectively. The increasing relative benefit at longer horizons reflects the rapid decay of the persistence assumption. Figure 6 visualizes the corresponding error comparison across horizons.

Table 5. Clean-input results.

Horizon	Model	MAE	RMSE	R^2
15	Persistence	0.0560	0.1072	0.805
15	Power-only ET	0.0426	0.0866	0.874
15	Median-impute ET	0.0376	0.0790	0.894
15	QDPCC-PV	0.0376	0.0790	0.894
60	Persistence	0.0853	0.1455	0.653
60	Power-only ET	0.0491	0.0955	0.848
60	Median-impute ET	0.0463	0.0926	0.857
60	QDPCC-PV	0.0463	0.0926	0.857
240	Persistence	0.1969	0.3000	-0.419
240	Power-only ET	0.0730	0.1365	0.711
240	Median-impute ET	0.0684	0.1261	0.753
240	QDPCC-PV	0.0684	0.1261	0.753

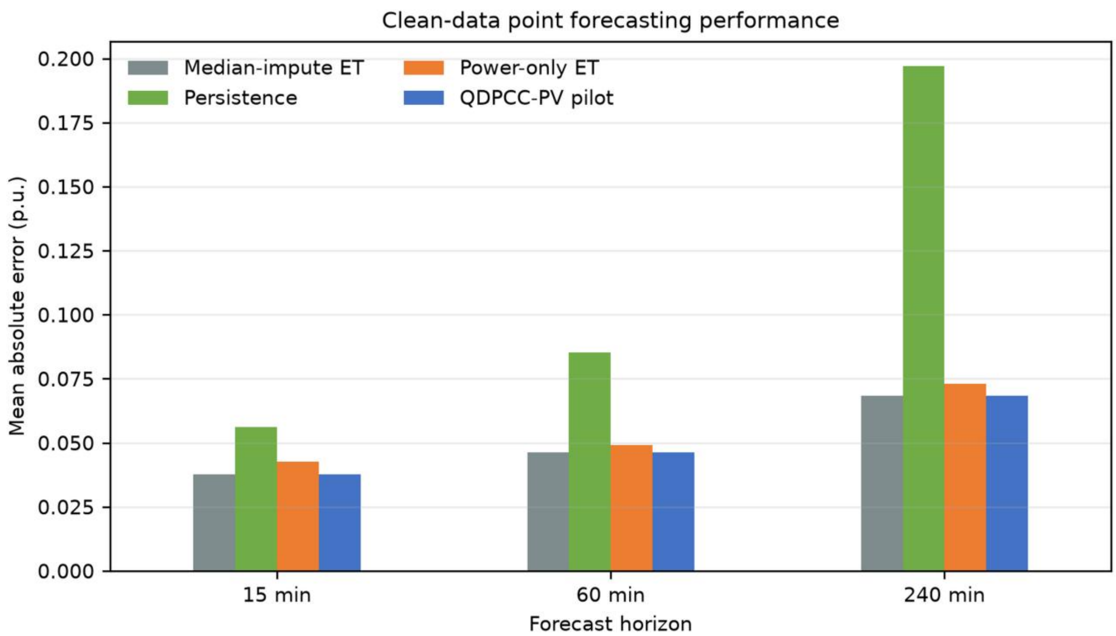

Figure 6. Clean-input forecasting errors.

The power-only expert remained competitive, confirming that lagged generation encodes recent cloud and plant behavior. The additional weather variables reduced MAE by 11.6% at 15 min, 5.7% at 60 min, and 6.3% at 240 min relative to power-only ET. These modest gains illustrate why a multivariate model should not be trusted unconditionally when the added variables are unreliable. Figure 7 provides a representative observed-versus-predicted sequence at the 60-min horizon.

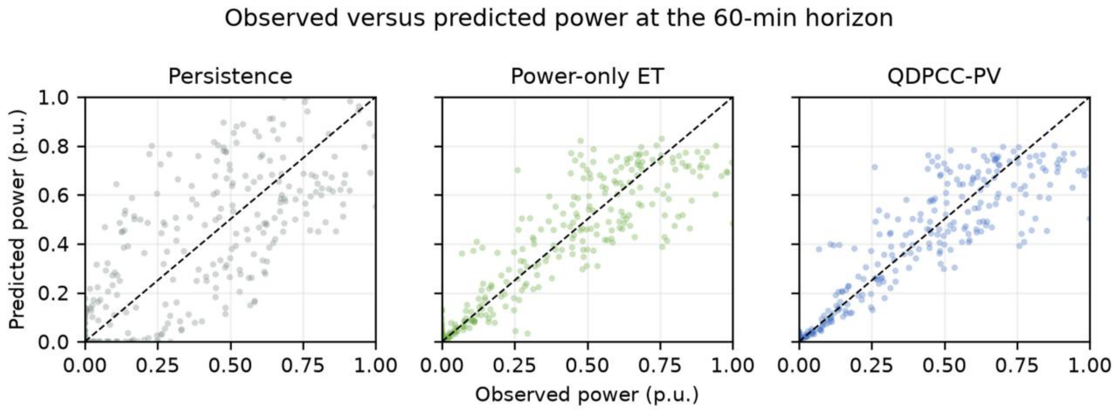

Figure 7. Observed and predicted power at 60 min.

6.2. Missing-Data Robustness

Figure 8 compares error as random meteorological missingness increases at the 60-min horizon. Persistence and power-only ET remain constant because neither receives weather variables. Median-impute ET degrades monotonically from 0.0463 MAE with complete inputs to 0.0787 under total weather loss. QDPCC-PV transfers weight toward the power-only expert and remains near 0.049 p.u. across the full range.

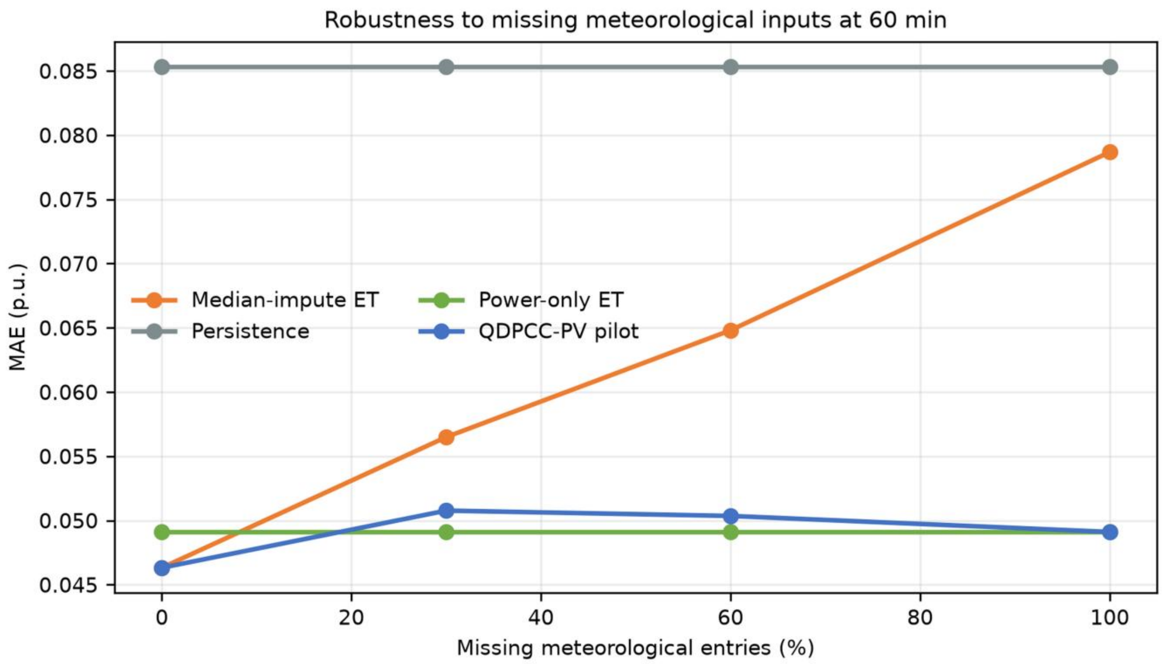

Figure 8. Error under weather-data loss.

Table 6. Results under random weather loss.

Missing	Model	MAE	RMSE
0%	Power-only ET	0.0491	0.0955
0%	Median-impute ET	0.0463	0.0926
0%	QDPCC-PV	0.0463	0.0926
30%	Power-only ET	0.0491	0.0955
30%	Median-impute ET	0.0565	0.1070
30%	QDPCC-PV	0.0508	0.0976
60%	Power-only ET	0.0491	0.0955
60%	Median-impute ET	0.0648	0.1214
60%	QDPCC-PV	0.0503	0.0963
100%	Power-only ET	0.0491	0.0955
100%	Median-impute ET	0.0787	0.1386
100%	QDPCC-PV	0.0491	0.0955

At 60% missingness, QDPCC-PV achieved 0.0492 MAE, whereas median-impute ET produced 0.0664. The 25.8% reduction supports the use of explicit quality routing. QDPCC-PV was approximately equal to the power-only expert at this point; the result is expected because only 40% of the weather entries remained available. Under total loss, both methods coincide by design.

The comparison also clarifies what the gate does and does not accomplish. It does not reconstruct irradiance, and it cannot make partially observed weather as informative as complete measurements. Instead, it limits the damage caused by presenting an imputed vector to a model trained on genuine observations. At intermediate missing rates, the fused prediction preserves part of the clean-input benefit while moving toward the stable fallback. Under complete loss, equivalence with the power-only expert is a required boundary check rather than a failure to improve. For block outages, the same logic operates over consecutive timestamps, avoiding the repeated use of a median-filled weather state throughout the gap. The remaining error is governed by the predictive content of recent power, daily periodicity, and the ability of the fallback expert to track ramps without direct weather information.

The scenario matrix is designed to separate robustness from clean-data accuracy. A model could achieve the smallest error with complete inputs yet fail sharply after a single weather channel disappears. Conversely, a power-only model can be stable but leave useful irradiance information unused. Reporting both clean and corrupted conditions reveals this trade-off. The median-imputation baseline is particularly informative because it uses the same weather-aware learner but lacks explicit routing; its deterioration therefore cannot be attributed to reduced tree capacity. The comparison indicates that numerical completeness is not equivalent to informational reliability. This observation is relevant beyond PV forecasting whenever auxiliary sensors are optional, delayed, or maintained by an external service.

6.3. Feature and Residual Analysis

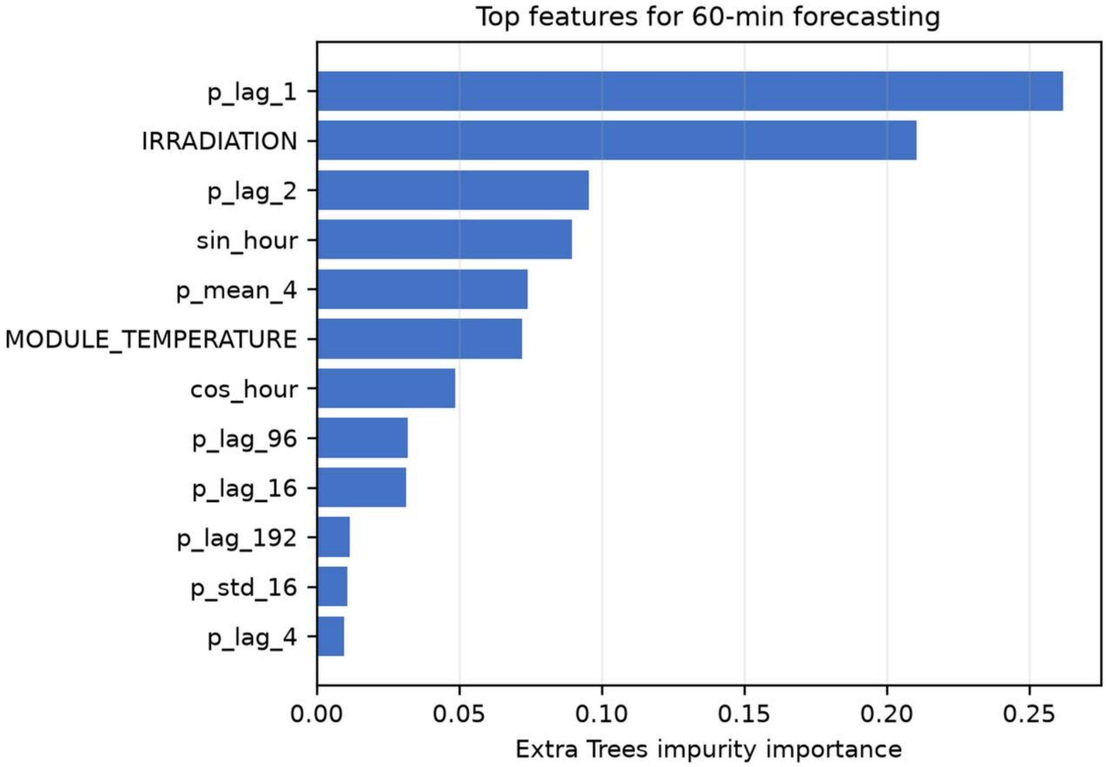

Figure 9. Extra Trees feature importance.

Figure 9 reports the Extra Trees importance ranking. The most important feature was the immediately preceding power value, followed by irradiance, the second power lag, time of day, and the one-hour rolling mean. Module temperature also contributed, whereas longer lags and short-window variability were less influential. Impurity importance is not causal and can be divided among correlated predictors, but the ordering is consistent with the correlation analysis and with the physical role of irradiance.

Figure 10 compares the residual distributions. All learned models reduced the central spread relative to persistence. The distributions remained heavy-tailed and slightly asymmetric because rapid cloud transitions were underrepresented. Median-impute ET and QDPCC-PV are identical under complete inputs, which provides a useful implementation check. Their behavior separates only after weather channels are removed.

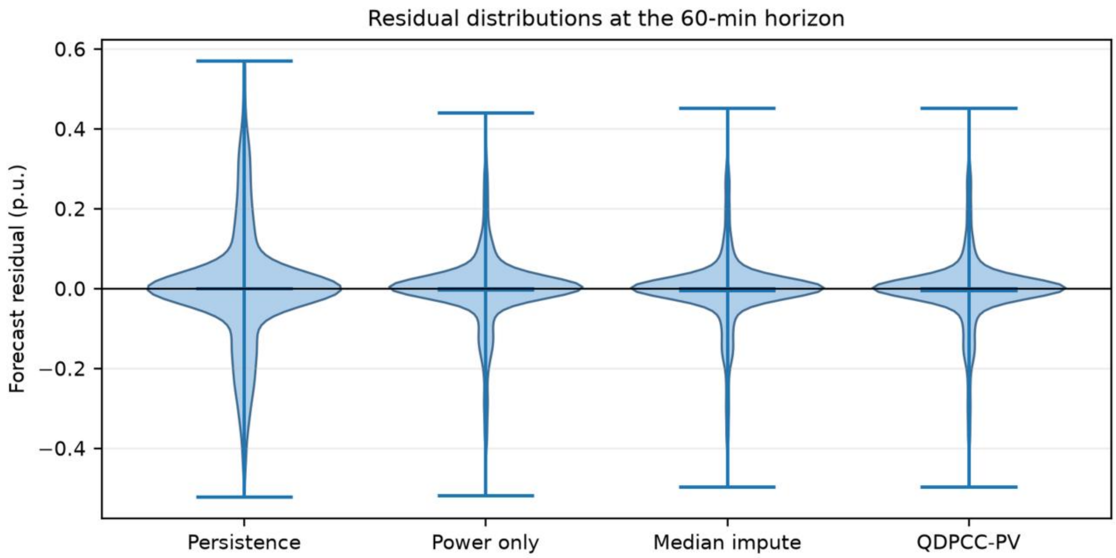

Figure 10. Residual distributions.

6.4. Interval Forecasts

Figure 11 shows a representative block-outage sequence. The gate moves to zero when all three weather channels are missing, activates the power-only expert, and returns to one when measurements resume. The point forecast follows the daily envelope but smooths abrupt intraday ramps. The conformal band expands according to the calibration residual quantile for the corresponding missingness condition.

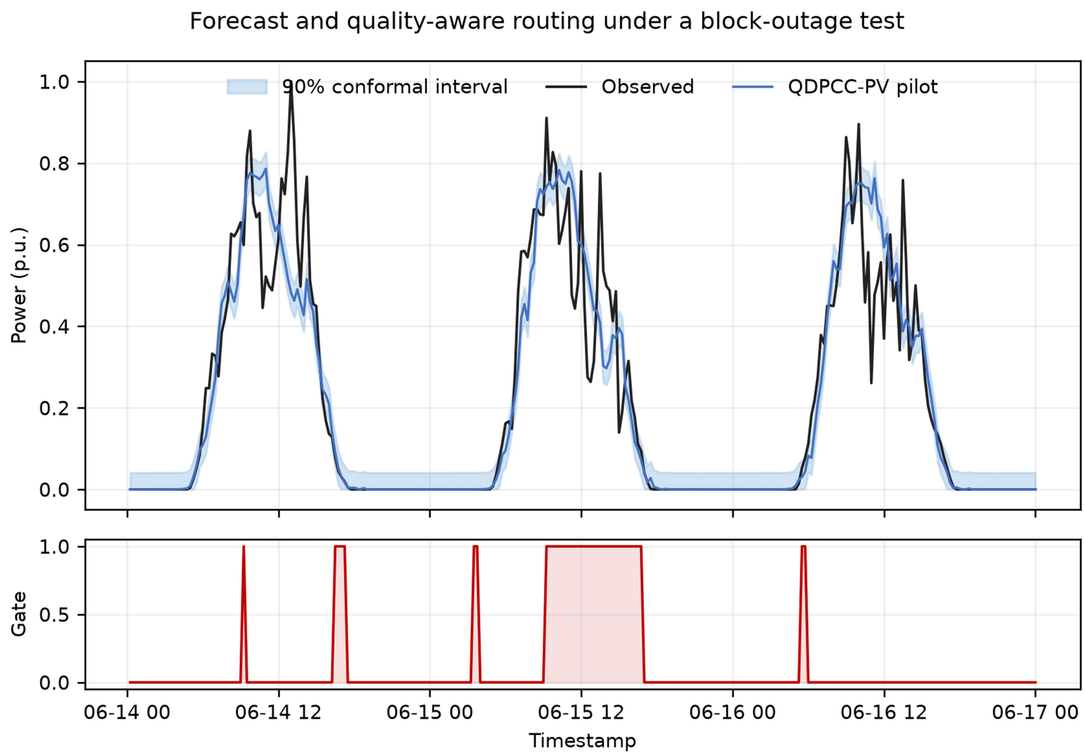

Figure 11. Forecasts during a weather-data outage.

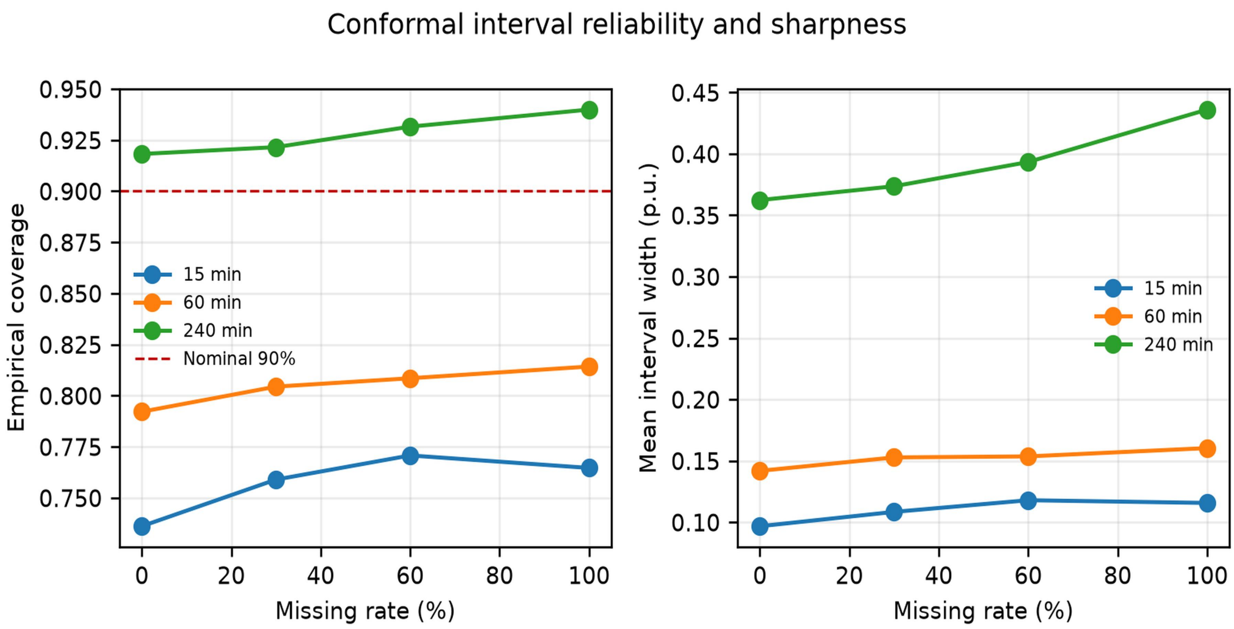

Figure 12. Interval coverage and width.

Table 7. Prediction-interval results.

Horizon	Missing	PICP	MPIW
15	0%	0.736	0.097
15	30%	0.759	0.108
15	60%	0.771	0.118
15	100%	0.765	0.116
60	0%	0.792	0.142
60	30%	0.804	0.153
60	60%	0.809	0.154
60	100%	0.814	0.160
240	0%	0.918	0.362
240	30%	0.922	0.374
240	60%	0.932	0.393
240	100%	0.940	0.436

Figure 12 and Table 7 summarize interval coverage and width across horizons and missingness levels. At 240 min, clean-input coverage was 0.918 and remained above 0.90 for all missing rates. Width increased from 0.362 to 0.436 as weather information disappeared. This is the desired qualitative behavior: uncertainty increases when input quality decreases. At 15 and 60 min, however, coverage remained below 0.90. Clean-input values were 0.736 and 0.792, and missingness-conditioned calibration raised them only to approximately 0.765 and 0.814 under total weather loss.

The undercoverage is not a numerical formatting issue; it is a substantive result. The final test period contains short-term ramps whose residual distribution differs from the preceding calibration block. A fixed split-conformal quantile cannot guarantee time-conditional coverage under this drift. Rolling calibration, weighted residuals, weather-regime Mondrian groups, or adaptive conformal inference should be evaluated in the extended study.

Coverage should also be interpreted at the decision horizon rather than only as a pooled average. A nominally calibrated interval may cover most nighttime zeros while missing a disproportionate number of daylight ramps. Conversely, widening all intervals to protect a small number of ramp events can reduce usefulness during stable periods. A follow-up study should stratify coverage by solar elevation, clear-sky index, ramp magnitude, and expert-gate state. It should report coverage deviation and width for each group and use rolling plots to identify when reliability changes. Such diagnostics would distinguish random finite-sample variation from systematic miscalibration caused by weather regimes or seasonal drift.

6.5. Practical Implications

The proposed framework has three deployment advantages. First, the power-only expert keeps producing forecasts during complete weather-service failure. Second, the gate itself is an interpretable data-quality indicator and can be logged for operator review. Third, conformal intervals expose uncertainty that is invisible in MAE and RMSE. The framework is modular: ET experts can be replaced by PVPNet, TCN, PatchTST, or another temporal model without changing the routing and calibration logic.

A practical implementation should place the feature builder, availability monitor, two experts, physical projection, and interval calibrator in separately testable components. The data-quality monitor should record channel presence, timestamp freshness, range violations, flat-line behavior, and disagreement with redundant measurements. Forecast logs should retain both expert outputs and the gate value so that operators can determine whether a change arose from the weather-aware path, the fallback path, or the routing decision. Alarm thresholds should be based on sustained deterioration rather than a single forecast error. Retraining and recalibration schedules should also be separated: the point models may remain stable for weeks, whereas residual quantiles may require more frequent updates when weather regimes shift. These controls convert the proposed equations into an auditable forecasting service rather than a one-time offline model.

Computationally, the dual-path design adds the cost of a second ET prediction, but both experts can run in parallel and the gate requires only a few arithmetic operations. The calibration stage stores residual quantiles rather than a generative distribution, so online interval construction is inexpensive. Memory and latency therefore remain compatible with plant-level supervisory systems. The most important operational risk is silent sensor degradation: a channel may be present but biased, stale, or physically implausible. Binary availability alone cannot detect that state. Future gates should combine missingness with freshness, range checks, cross-sensor consistency, and expert disagreement, while preserving a monotonic relation between lower input quality and lower reliance on the weather-aware expert.

The study also has important limits. The post-lag record covers 32 days, not a full annual cycle. Both plants are drawn from the same public collection and do not establish cross-climate generalization. Measured weather is used instead of genuine future NWP, so the reported values are best interpreted as short-horizon sensor-availability experiments. The deterministic gate depends only on the number of observed variables and does not distinguish the greater importance of irradiance. Finally, one model seed does not provide a formal significance test.

A complete follow-up experiment should use PVOD [2] or Desert Knowledge Australia Solar Centre data, retain at least one full year for seasonality, evaluate several stations, include genuine NWP issue times, run five or more random seeds, and compare with PVPNet, LSTM, TCN, PatchTST, and recent missingness-aware methods under identical splits. Statistical comparison should use Diebold-Mariano tests for forecast-loss series or paired Wilcoxon tests across site-seed units with Holm correction [26-28].

External validation should preserve the information available at each forecast issue time. Measured future irradiance must not be substituted for an operational weather forecast, and station metadata should be frozen before test evaluation. Site transfer should distinguish zero-shot application from local fine-tuning, while seasonal transfer should report the amount of target-season data used. Computational comparisons should include training time, peak memory, parameter count, and inference latency in addition to error. These controls are necessary for determining whether a more complex deep expert produces a material operational benefit over the transparent ET implementation. They also prevent gains caused by different preprocessing, target scaling, or data leakage from being attributed incorrectly to the forecasting architecture.

7. Conclusion

This paper introduced QDPCC-PV, a quality-aware dual-path framework for photovoltaic power forecasting when meteorological variables may be incomplete. A multivariate ET expert exploits irradiance and temperature during normal operation, while a power-only ET expert provides a stable fallback. The observed-feature ratio determines the convex fusion weight, a physical projection enforces feasible normalized output, and missingness-conditioned conformal calibration provides interval forecasts.

Experiments on two measured PV plants showed that the proposed point forecast outperformed persistence at 15-, 60-, and 240-min horizons. Under 60% weather missingness at the 60-min horizon, it reduced MAE from 0.0664 for median-imputation forecasting to 0.0492. The dual path therefore prevented a major loss of accuracy when covariates were corrupted. The interval study produced a more qualified conclusion: 240-min coverage met the nominal 90% level, but 15- and 60-min intervals undercovered because of temporal residual drift.

The main conclusion is that input availability should be represented explicitly in operational PV forecasting. Imputation can make an input numerically complete without making it trustworthy. Future work will replace the deterministic gate with a learned quality model, evaluate CNN and PatchTST experts on multi-year multi-station data, incorporate NWP age and weather regimes, and use adaptive conformal inference to maintain short-horizon coverage.

Appendix A. Feature Definitions

Table A1. Input features.

Feature	Description	Expert
p_lag_1 ... p_lag_192	Power at selected prior steps	Both
p_mean_4,16,96	Shifted rolling mean	Both
p_std_4,16,96	Shifted rolling standard deviation	Both
sin_hour, cos_hour	Cyclic time-of-day encoding	Both
sin_doy, cos_doy	Cyclic day-of-year encoding	Both
IRRADIATION	Plant-level solar irradiance	Multivariate
AMBIENT_TEMPERATURE	Ambient air temperature	Multivariate
MODULE_TEMPERATURE	PV module temperature	Multivariate
m_t,j	Observed/missing weather mask	Gate

Appendix B. Reproducibility Checklist

The experiment is reproducible from the raw generation and weather comma-separated-value files using the following sequence: parse timestamps; sum inverter AC power; average plant weather; construct the 15-min index; normalize power by the training-safe capacity proxy; generate shifted lags and rolling features; split chronologically; fit training medians; train both ET experts; generate calibration and test missingness with seed 20260917; compute gate weights and fused predictions; calibrate condition-specific residual quantiles; and calculate point and interval metrics. The final implementation should record software versions, processor information, wall-clock training time, and dataset checksums.

References

[1] Huang, C.-J., & Kuo, P.-H. (2019). Multiple-input deep convolutional neural network model for short-term photovoltaic power forecasting. IEEE Access, 7, 74822-74834. https://doi.org/10.1109/ACCESS.2019.2921238

[2] Yao, T., Wang, J., Wu, H., Zhang, P., Li, S., Wang, Y., Chi, X., & Shi, M. (2021). A photovoltaic power output dataset: Multi-source photovoltaic power output dataset with Python toolkit. Solar Energy, 230, 122-130. https://doi.org/10.1016/j.solener.2021.09.050

[3] Holmgren, W. F., Hansen, C. W., & Mikofski, M. A. (2018). pvlib python: A Python package for modeling solar energy systems. Journal of Open Source Software, 3(29), 884.

[4] Das, U. K., Tey, K. S., Seyedmahmoudian, M., Mekhilef, S., Idris, M. Y. I., Van Deventer, W., Horan, B., & Stojcevski, A. (2018). Forecasting of photovoltaic power generation and model optimization: A review. Renewable and Sustainable Energy Reviews, 81, 912-928. https://doi.org/10.1016/j.rser.2017.08.017

[5] Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with Transformers. In Proceedings of the International Conference on Learning Representations.

[6] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (Vol. 30).

[7] Lim, B., Arik, S. O., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748-1764.

[8] Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv:1803.01271.

[9] [9] Abdel-Nasser, M., & Mahmoud, K. (2019). Accurate photovoltaic power forecasting models using deep LSTM-RNN. Neural Computing and Applications, 31, 2727-2740.

[10] [Gensler, A., Henze, J., Sick, B., & Raabe, N. (2016). Deep learning for solar power forecasting using AutoEncoder and LSTM neural networks. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (pp. 2858-2865).

[11] El-Kenawy, E.-S. M., Alhussan, A. A., Mattar, E. A., & Radwan, M. (2026). Feature selection and hyperparameter tuning in transformer-based deep learning models for photovoltaic power forecasting using the Swordfish Movement Optimization Algorithm. International Journal of Electrical Power & Energy Systems, 174, 111509. https://doi.org/10.1016/j.ijepes.2025.111509

[12] Tao, K., Zhao, J., Tao, Y., Qi, Q., & Tian, Y. (2024). Operational day-ahead photovoltaic power forecasting based on transformer variant. Applied Energy, 373, 123825. https://doi.org/10.1016/j.apenergy.2024.123825

[13] Sun, Q., Yan, F., Sun, W., & Zhou, Y. (2025). DWT-Former: Fusing wavelet-based multi-scale features and transformer-based temporal representations for photovoltaic power forecasting. Energy, 341, 139283. https://doi.org/10.1016/j.energy.2025.139283

[14] Liu, Y., & Wang, Z. (2025). Robust and efficient multihorizon photovoltaic power forecasting with a dilated multi-scale Transformer. Solar Energy, 302, 114077. https://doi.org/10.1016/j.solener.2025.114077

[15] Yan, K., Liu, J., Zhang, J., Yang, F., Gao, Y., & Du, Y. (2025). Robust photovoltaic forecasting under severe data missingness via multi-domain collaboration and covariate interaction. Applied Energy, 401, 126771. https://doi.org/10.1016/j.apenergy.2025.126771

[16] Vovk, V., Gammerman, A., & Shafer, G. (2005). Algorithmic learning in a random world. Springer.
[17] Romano, Y., Patterson, E., & Candès, E. J. (2019). Conformalized quantile regression. In Advances in Neural Information Processing Systems (Vol. 32).
[18] Xu, C., & Xie, Y. (2021). Conformal prediction interval for dynamic time-series. In Proceedings of the 38th International Conference on Machine Learning (pp. 11559-11569).
[19] Gibbs, I., & Candès, E. (2021). Adaptive conformal inference under distribution shift. In Advances in Neural Information Processing Systems (Vol. 34).
[20] Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32.
[21] Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine Learning, 63, 3-42.
[22] Liu, M., Lai, Z., Fang, Y., & Ling, Q. (2025). Day-ahead photovoltaic power forecasting based on corrected numeric weather prediction and domain generalization. Energy and Buildings, 329, 115212. https://doi.org/10.1016/j.enbuild.2024.115212
[23] Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., & Hyndman, R. J. (2016). Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. International Journal of Forecasting, 32(3), 896-913.
[24] Renkema, Y., Visser, L., & AlSkaif, T. (2024). Enhancing the reliability of probabilistic PV power forecasts using conformal prediction. Solar Energy Advances, 4, 100059. https://doi.org/10.1016/j.seja.2024.100059
[25] Solar Power Generation Data. (2020). Kaggle. https://www.kaggle.com/datasets/anikannal/solar-power-generation-data
[26] Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253-263.
[27] Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics Bulletin, 1(6), 80-83.
[28] Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6(2), 65-70.