

Article

A Validation-Guided Framework for Target-Specific Pathway Configuration in Daily Electricity-Supply Forecasting Under Data Constraints

Lian Tan * , Yande Chen, Pengfei Hou , Yaming Li, Lanting Zeng, Zifan Yang and Tongchui Liu 

School of Electrical Engineering, North China University of Water Resources and Electric Power, No. 36 Beihuan Road, Jinshui District, Zhengzhou 450045, China; 13013880022@163.com (Y.C.); 15238349569@163.com (P.H.); liyaming_mail@163.com (Y.L.); lvntin@163.com (L.Z.); yangzifan1218@163.com (Z.Y.); lbhammer@163.com (T.L.)

* Correspondence: tanunited@ncwu.edu.cn

Abstract

Accurate daily forecasts of thermal generation, renewable generation, hydropower, and imported electricity are essential for spot-market trading and day-ahead scheduling. However, short training histories, incomplete operational information, and distinct target drivers constrain accuracy and generalizability. To address these challenges, this study proposes a Delta-enhanced, validation-guided framework for configuring target-specific forecasting pathways. The Delta representation captures inter-day changes in thermal and hydropower generation. For the internal renewable-generation target, a regional target-day weather-forecast interface is applied retrospectively. It aggregates 13 hierarchical regional forecast columns for each of 24 weather variables and incorporates six calendar variables. For each target, chronological validation selects an admissible pathway defined by forecast-origin information, target representation, predictor family, optional temporal representation, and output rule. Once selected, it is refitted and fixed for subsequent forecasting. The evaluation covers four internal targets from a regional power system in China and six national-scale targets from France. The 97-day evaluation yielded R^2 values of 0.8084, 0.7337, 0.7782, and 0.5212 for thermal generation, renewable generation, hydropower, and imported electricity, respectively. The frozen thermal and renewable pathways reduced RMSE by 23.2% and 50.9%, respectively, relative to Persistence, a baseline that carries the previous-day value forward. These findings support target-specific forecasting under data constraints.

Keywords: daily electricity-supply forecasting; sparse data; incomplete operational information; target-specific pathway configuration; chronological validation; Delta representation; capacity-normalized representation; electricity market operation



Academic Editor: Seung-Hoon Yoo

Received: 4 August 2026

Revised: 5 September 2026

Accepted: 7 September 2026

Published: 8 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Day-ahead forecasts of heterogeneous electricity-supply components support spot-market bidding, adjustments to contracted electricity volumes, generation scheduling, reserve allocation, and interregional electricity procurement. In regional power systems, the relevant supply portfolio typically comprises conventional thermal generation, variable renewable generation, hydropower, and imported electricity [1,2]. Each component has distinct operational drivers. Thermal output depends on unit-commitment schedules, operating status, and fuel-supply constraints [3]. Renewable output is driven mainly by meteorological variability and spatial complementarity [4–6], whereas hydropower output depends on streamflow, reservoir states, and dispatch decisions [7]. Imported electricity is

also shaped by transmission arrangements, transaction schedules, and decisions outside the receiving region [8–11]. These differences make it difficult for a single forecasting configuration to achieve consistently high accuracy across the portfolio. The challenge is compounded by short daily training histories and incomplete operational information. Existing studies have therefore explored ways to organize historical, meteorological, and source-specific information and combine it with models tailored to individual generation sources.

Existing renewable and multi-source forecasting studies have improved input construction by integrating historical generation, meteorological variables, and source-specific descriptors [12]. Building on these multi-source inputs, climate-feature analysis and enhanced tree models have been used for medium- and long-term photovoltaic forecasting [13]. Decomposition and similar-day selection have also been combined with recurrent networks for day-ahead prediction [14]. At the regional scale, grid partitioning and ensemble learning have also been adopted to represent spatial organization and source composition in photovoltaic, wind, and aggregated renewable forecasting [15]. More recent studies have extended this direction to coordinated and dependency-aware forecasting. In integrated energy systems, shared representations have been used to jointly forecast electricity, cooling, and heating loads by exploiting inter-energy dependencies [16]. For geographically distributed renewable generation, graph-based spatio-temporal models combine the learning of cross-site spatial dependencies with temporal feature extraction [17]. At the system level, ensemble approaches integrate complementary machine-learning predictors to forecast aggregated wind and solar generation [18]. From a market perspective, cross-border electricity trading has also become an important component of regional power-system coordination, particularly in systems with substantial hydropower and renewable-energy resources [19]. These studies demonstrate that forecasting accuracy depends not only on the amount of information available but also on how temporal, meteorological, and spatial information is organized. However, most studies focus on a single generation source or a fixed aggregation level. Their information interfaces are generally predefined rather than selected to reflect the operational drivers and information constraints of individual electricity-supply targets.

Recent studies have applied recurrent-network-XGBoost hybrids, Transformer- or BiLSTM-enhanced convolutional models, boosted trees, and multi-task architectures to renewable-energy forecasting [20–26]. Although these methods enhance nonlinear learning and temporal representation, greater architectural complexity does not necessarily yield more reliable forecasts when daily records are limited. Moreover, many model comparisons vary only the predictor while holding the input interface and target representation fixed. For heterogeneous electricity-supply forecasting, predictor families and temporal representations should therefore be configured for each target rather than assumed to be universally optimal.

Other studies have improved individual stages of the forecasting process. Weak stationarization and decomposition have been used to address time-series non-stationarity [27,28]. Mixture-of-experts models and dynamic ensembles adjust model contributions across temporal scales and operating regimes [29,30]. Explainable tree models identify influential variables, whereas conformal and residual-based intervals characterize uncertainty around point forecasts [31–34]. These developments improve specific components of the forecasting process but generally assume that the information interface and model structure have already been specified. Conventional time-series pipeline selection and automated machine learning, or AutoML, provide systematic mechanisms for configuring forecasting workflows within predefined search spaces. Recent reviews describe automated forecasting pipelines as coordinated combinations of data preprocessing, feature engineer-

ing, hyperparameter optimization, forecasting-method selection, and ensembling [35,36]. Energy-oriented frameworks also integrate feature selection, model selection, ensembling, and postprocessing within configurable workflows [37]. Related studies have explored dynamic ensemble selection from heterogeneous forecasting pools [38] and multi-stage AutoML approaches that automate feature processing, model generation, and model selection [39]. Together, these studies show that pipeline-level methods can coordinate multiple stages of a forecasting workflow. However, heterogeneous electricity-supply forecasting requires an additional configuration layer that accounts for information availability at the forecast origin and the admissibility of target-specific representations. For this task, each target has a distinct set of admissible information sources and target representations at the forecast origin. These constraints define the target's pathway configuration space. Accordingly, the configuration scheme defines an admissible pathway space for each target under explicit information and temporal constraints before candidate pathways are compared. This separation is particularly important when weather information, operating plans, and installed-capacity records have different availability boundaries.

To address this gap, this study develops a validation-guided, target-specific forecasting pathway-configuration methodology for heterogeneous electricity-supply forecasting under explicit forecast-origin information constraints. The complete forecasting pathway serves as the unit of configuration and jointly specifies the admissible information interface, target representation, predictor family, optional temporal representation, and output rule. For each target, pathway eligibility is established before validation according to target-specific physical relevance and forecast-origin data availability, thereby defining a target-specific admissible pathway space. Chronological validation is then used to select among the admissible pathways. The selected pathway is subsequently refitted using admissible pre-evaluation data, frozen, and used unchanged in subsequent forecasting. Under this procedure, thermal generation retains a plan-informed XGBoost–Delta pathway, whereas renewable generation retains an F2-Regional Random-Forest Level pathway. Hydropower retains a PCA-enhanced Transformer–XGBoost Delta pathway with a persistence-based output guard, whereas imported electricity retains Persistence. The framework is evaluated on four internal targets in a Chinese regional power system and six national-scale targets from France. Additional analyses quantify performance changes under shorter training histories and unavailable key inputs. The temporal availability boundary of each weather interface is also stated explicitly. Accordingly, the methodological contribution centers on target-specific forecasting pathway configuration and its chronological protocol. Established forecasting models serve as constituent predictors, while operational applications provide the deployment context. The main contributions are summarized as follows:

- (1) A validation-guided, target-specific forecasting pathway-configuration methodology is developed for heterogeneous electricity-supply forecasting under explicit forecast-origin information constraints. Each candidate pathway jointly specifies the admissible information interface, target representation, predictor family, optional temporal representation, and output rule, while target-specific eligibility is determined before validation according to physical relevance and data availability at the forecast origin. This design enables each forecasting target to be configured within its own admissible candidate pathway space.
- (2) Target-specificity is implemented through the joint organization of information interfaces and target representations within the admissible pathway space. Level, Delta, and previous-year-capacity-normalized 24 h utilization are retained as eligible representation choices according to the physical characteristics and forecast-origin data availability of each target. The information interface is likewise tailored to target-specific operational drivers: plan-informed operating context is available to thermal

and hydropower candidates, while the F2-Regional interface organizes hierarchical regional forecast information and calendar variables for the internal regional renewable-generation target. Chronological validation then configures these target-specific information and representation components jointly with the predictor.

- (3) An auditable chronological configuration and evaluation protocol organizes eligibility screening, validation, pathway selection, refitting, freezing, and fixed subsequent execution. Fully subsequent internal confirmation, rolling-origin external evaluation, controlled training-history reduction, information-unavailability scenarios, and residual-based uncertainty analysis are assigned distinct chronological roles. This structure enables separate assessment of out-of-sample generalization, training-history sensitivity, information dependence, and predictive uncertainty while preserving the pathway decisions established during configuration. Controlled contemporary-baseline comparisons, factorial ablation, and repeated-window stability analyses further characterize the target-dependent value and stability of pathway components.

2. Methodology

2.1. Problem Formulation for Heterogeneous Targets

The internal study uses data from a regional electricity system in China, and comprises four heterogeneous day-ahead electricity-volume targets: plant-level thermal generation (TH), regional renewable generation (RE), regional hydropower generation (HY), and regional imported electricity (IM). The external analysis uses French RTE data and considers six national-scale targets: fossil thermal generation, renewable generation, hydropower generation, net physical exchange, solar generation, and wind generation:

$$\mathcal{J} = \{\text{TH, RE, HY, IM}\} \quad (1)$$

where $\mathcal{J}$ denotes the target set. TH is defined at the plant level, whereas RE, HY, and IM are regional electricity-supply components. In this study, electricity-supply forecasting covers both electricity generated within the regional system and electricity imported from outside the region, whereas generation forecasting refers only to generation-side output. Accordingly, TH, RE, and HY are generation-side supply components, while IM represents an external supply component. Imported electricity and external net physical exchange have different physical meanings and are therefore treated as separate forecasting targets in this study.

For target j in $\mathcal{J}$ and target day d , the one-day-ahead forecast is expressed as follows:

$$\hat{y}_{j,d} = F_j(x_{j,d}), j \in \mathcal{J} \quad (2)$$

where the symbols are defined above. Historical lags, rolling statistics, and Delta reconstruction anchors use observations available through $D - 1$. Target-day calendar variables are treated as available at the forecast stage. Weather timing is interface-specific: F2-Regional is retrospectively assembled from archived previous-run forecasts for target day D with an approximately one-day nominal lead; F2-RTE-Hist uses lagged and rolling historical weather available before the external forecast origin; and W2-RTE uses retrospective target-day ERA5 weather in the designated retrospective complete-weather benchmark.

The information set is specified separately for each target. Thermal candidates combine historical and calendar variables with generation-plan information. Internal renewable candidates use 24 target-day weather-forecast variables and six calendar variables. Hydropower candidates may include historical variables, ProvincePlanT+Lag, and latent temporal features, whereas imported-electricity candidates use historical and calendar information.

ProvincePlanT+Lag denotes a plan-informed operating context that combines a province-level target-day generation-plan term available at the forecast origin with lagged, rolling, interaction, and historical-change features derived from earlier plan records. It is eligible for thermal and hydropower candidates. Predicted-state coupling, denoted PSC, is a 13-dimensional chronological predicted-state interface constructed from forecast-origin-valid upstream prediction states and their derived changes, shares, and balance descriptors. PSC is used only as an alternative renewable-information pathway, and realized target-day generation values are excluded.

The component-wise constructions of ProvincePlanT+Lag and PSC are specified in Sections 2.5.1 and 2.5.2, respectively.

Candidate pathways combine the available information with Level, Delta, or capacity-normalized representations and eligible predictors. Chronological validation then compares the admissible pathways for each target.

2.2. Overall Framework

The procedure comprises four steps. The complete forecasting pathway is the configuration unit, and its admissible space is defined separately for each target before validation. First, physically relevant inputs available at the forecast stage define the candidate set for each target. Second, the candidate pathways are compared by chronological validation. Third, model parameters are refitted using the available pre-evaluation data after the pathway has been selected. Finally, the selected configuration is frozen and applied to the corresponding evaluation fold or subsequent confirmation period. The same sequence is used for all targets, while the candidate set differs by target.

The overall architecture of the proposed framework is illustrated in Figure 1. Figure 1a presents the construction of target-specific candidate pathways, where each pathway is defined by the combination of information interface, target representation, predictor component, temporal representation, and output rule. Figure 1b presents the chronological validation and fixed execution procedure, separating offline configuration from subsequent frozen forecasting.

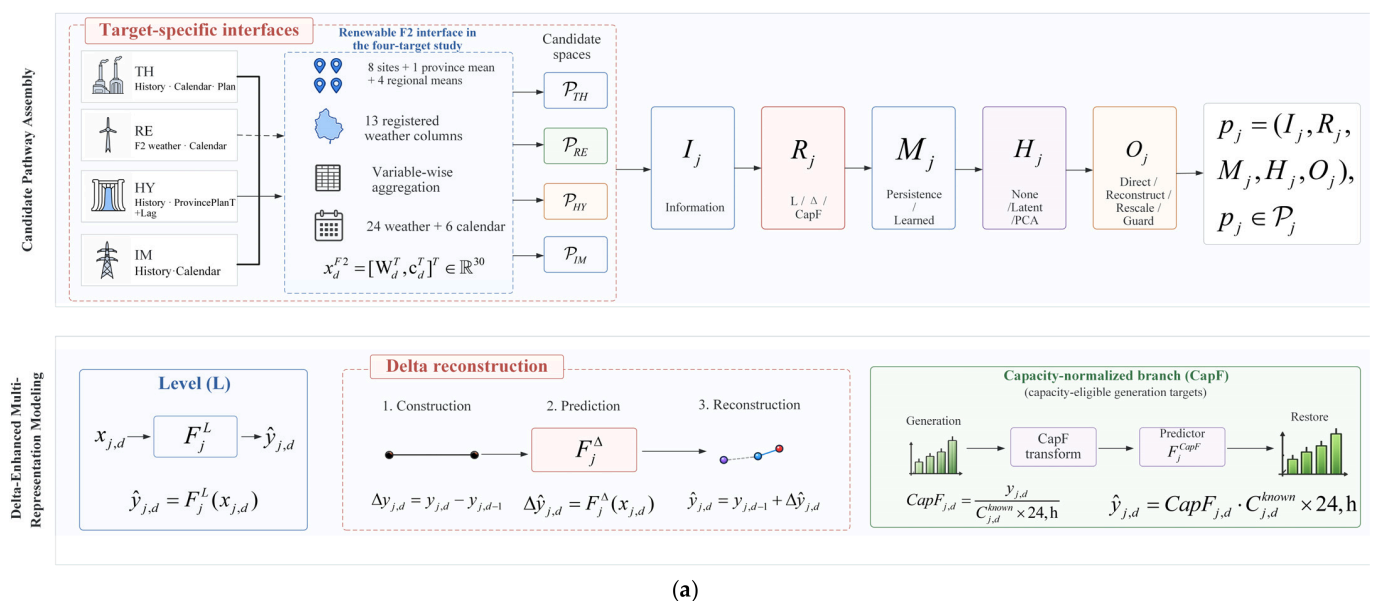


Figure 1. Cont.

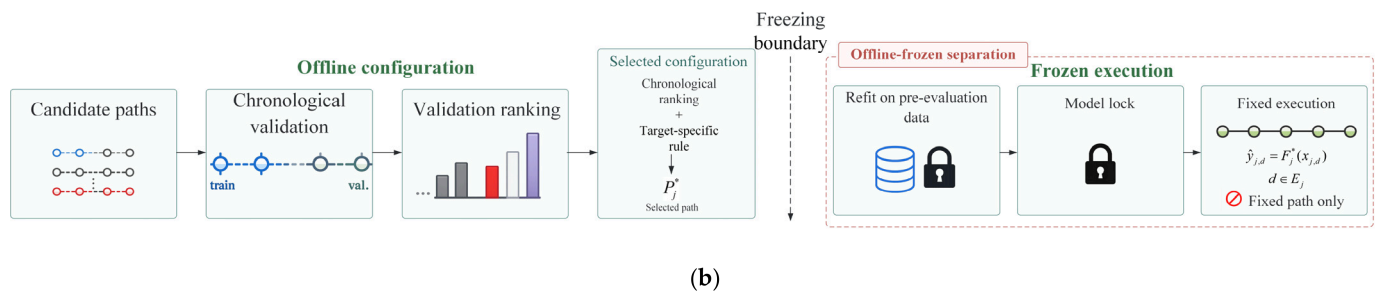


Figure 1. Overview of the proposed target-specific pathway configuration framework. (a) Target-specific candidate pathway assembly, including information interfaces, target representations, predictor components, and output rules. (b) Chronological validation and frozen execution procedure, including candidate pathway evaluation, pathway selection, refitting, model locking, and fixed execution. The colors distinguish schematic pathways and functional components; they do not encode quantitative values. In the lower part of (a), blue, red, and green identify the Level, Delta, and capacity-normalized branches, respectively. TH, RE, HY, and IM denote thermal generation, renewable generation, hydropower, and imported electricity. The components I_j , R_j , M_j , H_j , and O_j represent the information interface, target representation, predictor, temporal representation, and output rule, respectively. Arrows indicate information flow or processing order, and dashed outlines group related components. In (b), the vertical dashed line marks the freezing boundary. The database icon represents the pre-evaluation data used for refitting, the lock icons indicate fixed configuration or model status, and the prohibition symbol emphasizes that only the frozen pathway is used during subsequent execution. The nodes and bars are schematic illustrations rather than measured data. Please include these explanations in the figure caption.

2.3. Level, Delta, and Capacity-Normalized Target Representations

Target representation is treated as a configurable element because the four targets differ in level persistence, variation amplitude, and sensitivity to incomplete operating drivers. In a Level pathway, the predictor directly estimates the absolute electricity volume from the information available at the forecast origin. This form preserves the original scale and is appropriate when the target level itself provides the dominant predictive signal.

In a Delta pathway, the inter-day variation of target j on day d is defined as follows:

$$\Delta y_{j,d} = y_{j,d} - y_{j,d-1} \quad (3)$$

where $\Delta y_{j,d}$ denotes the inter-day change of target j on day d ; $y_{j,d}$ and $y_{j,d-1}$ are the observed electricity volumes of target j on days d and $d - 1$, respectively; j indexes the forecasting target; and d indexes the forecast date.

The corresponding predictor estimates the next-day variation within the same forecast-origin information boundary:

$$\widehat{\Delta y}_{j,d} = F_j^{\Delta}(x_{j,d}) \quad (4)$$

where $\widehat{\Delta y}_{j,d}$ denotes the predicted inter-day variation of target j on day d ; $F_j^{\Delta}(\cdot)$ is the target-specific forecasting function under the Delta representation; $x_{j,d}$ is the input vector available for target j at the forecast origin on day d ; and j and d index the forecasting target and date.

The day-ahead electricity-volume forecast is then reconstructed on the original scale:

$$\hat{y}_{j,d} = y_{j,d-1} + \widehat{\Delta y}_{j,d} \quad (5)$$

where $\hat{y}_{j,d}$ is the reconstructed one-day-ahead forecast of target j on day d ; $y_{j,d-1}$ is the latest observed electricity volume available at the forecast origin; and $\widehat{\Delta y}_{j,d}$ is the predicted inter-day variation obtained from Equation (4).

Level and Delta are evaluated in parallel. The selected representation is determined together with the information set and predictor during chronological configuration.

For external generation targets with available installed-capacity information, CapF is retained as a compact pathway label for previous-year-capacity-normalized 24 h daily utilization:

$$\text{CapF}_{j,d} = \frac{y_{j,d}}{C_{j,d}^{\text{known}} \times 24 \text{ h}} \quad (6)$$

where the denominator uses the forecast-origin-safe installed capacity registered at the end of year $y - 1$ for predictions in year y and a fixed 24 h convention. Renewable capacity is the sum of registered wind and solar capacity. CapF is applied to targets with an associated installed-capacity series, while net physical exchange is modeled using the Level and Delta representations. The fixed 24 h denominator is used as a common normalization convention for the CapF representation across all calendar dates. The daily-energy numerator retains the actual energy accumulated over each Europe/Paris local day, including the 23 h and 25 h daylight-saving transition days. Thus, the fixed denominator standardizes the capacity-normalized representation while preserving the actual daily energy in the numerator. For the external factorial analysis, capacity scaling (Raw/CapF) and change representation (Level/Delta) are treated as orthogonal subdimensions within the broader target-representation dimension.

2.4. Candidate Predictors and Representation Components

The internal four-target candidate set comprises Persistence, Random Forest (RF), XGBoost, LightGBM, and optional Transformer-PCA representation enhancement. The external candidate space includes Native, historical-weather, and generic capacity-normalized inputs combined with Ridge, RF, XGBoost, LightGBM, or Persistence when applicable. For renewable, solar, and wind generation, the retrospective target-day weather family contains four capacity-normalized XGBoost or TXGB candidates (C1–C4):

$$\hat{y}_{j,d}^{\text{Per}} = y_{j,d-1} \quad (7)$$

where $\hat{y}_{j,d}^{\text{Per}}$ denotes the Persistence forecast for target j on day d ; $y_{j,d-1}$ is the most recently observed electricity volume; and the superscript Per identifies the Persistence predictor.

RF and XGBoost serve as the principal nonlinear predictors. RF reduces variance by averaging decorrelated regression trees constructed from bootstrap samples and randomly selected feature subsets. XGBoost builds trees sequentially to correct preceding residuals and uses explicit regularization to control complexity [40]. Their outputs can be expressed in a common form:

$$s_{j,d}^{(m)} = f_m(\mathbf{u}_{j,d}), m \in \{\text{RF}, \text{XGB}\} \quad (8)$$

where $s_{j,d}^{(m)}$ denotes the output of predictor family m for target j on day d ; $f_m(\cdot)$ is the regression function associated with model family m ; $\mathbf{u}_{j,d}$ is the corresponding model-input vector, which may contain only admissible original variables or may be augmented by latent temporal features; $m \in \{\text{RF}, \text{XGB}\}$ identifies Random Forest or XGBoost, respectively; and j and d index the forecasting target and date.

For candidates requiring additional temporal representation, a Transformer encoder maps the historical input window to a sequence of hidden states [41,42]. Consistent with the implemented branch, temporal average pooling followed by LayerNorm produces a fixed-length latent vector:

$$\mathbf{H}_{j,d} = E_\theta(\mathbf{X}_{j,d-L:d-1}), \mathbf{h}_{j,d} = \text{LayerNorm}[\text{Mean}(\mathbf{H}_{j,d})] \quad (9)$$

where $X_{j,d-L:d-1}$ is the historical input sequence of target j from day $d - L$ to day $d - 1$; L is the lookback-window length; $E_\theta(\cdot)$ is the Transformer encoder parameterized by θ ; $H_{j,d}$ is the resulting sequence of hidden states; $\text{Mean}(\cdot)$ denotes temporal average pooling; $\text{LayerNorm}(\cdot)$ denotes layer normalization; and $h_{j,d}$ is the resulting fixed-length latent vector.

PCA may then compress the training-standardized latent vector, after which the compressed representation is concatenated with the admissible forecast-origin variables before tree-based regression:

$$z_{j,d}^{(k)} = W_k^T D_\sigma^{-1} (h_{j,d} - \mu_h), \tilde{x}_{j,d} = [x_{j,d}; z_{j,d}^{(k)}] \quad (10)$$

where $z_{j,d}^{(k)}$ is the k -dimensional PCA-compressed latent vector; W_k is the loading matrix formed by the first k principal components estimated from the training data; the superscript T denotes matrix transpose; μ_h and D_σ are, respectively, the training-set mean vector and diagonal standard-deviation scaling matrix of the latent features; $h_{j,d}$ is the latent vector defined in Equation (9); $x_{j,d}$ is the admissible original input vector; $\tilde{x}_{j,d}$ is the augmented tree-model input; and $[\cdot; \cdot]$ denotes vector concatenation.

The complete latent vector and its PCA-compressed form are both optional representation components. Accordingly, the tree-model input can be formed from the original variables alone, the original variables plus the complete latent vector, or the original variables plus the compressed latent vector.

For hydropower, a persistence-based guard is applied after Delta reconstruction. The guard addresses physically infeasible reconstructed outputs and is triggered when the raw forecast is negative or non-finite:

$$\hat{y}_{j,d}^{out} = (1 - g_{j,d}) \hat{y}_{j,d}^{model} + g_{j,d} \hat{y}_{j,d}^{Per}, g_{j,d} \in \{0, 1\} \quad (11)$$

where $\hat{y}_{j,d}^{out}$ is the final guarded forecast; $\hat{y}_{j,d}^{model}$ is the model output before guarding; $\hat{y}_{j,d}^{Per}$ is the Persistence forecast defined in Equation (7); and $g_{j,d} \in \{0, 1\}$ is described as follows: the guard takes $g_{j,d} = 1$ for a negative or non-finite raw forecast and replaces it with $\hat{y}_{j,d}^{Per}$; otherwise $g_{j,d} = 0$.

TXGB denotes a Transformer/PCA temporal representation followed by XGBoost regression. Together with RF, XGBoost, Persistence, and the optional output guard, these components form the target-specific pathways in the four-target study. In the external study, TXGB is used in the retrospective target-day weather candidates C3 and C4.

To further evaluate the forecasting performance of the frozen pathways during the fully subsequent evaluation stage, TimeXer [43] and PatchTST [44] were introduced as additional contemporary Transformer-based forecasting baselines. TimeXer models the target series jointly with exogenous variables to represent the contribution of external information, whereas PatchTST uses patch-based sequence representations to capture temporal dependencies. Both baselines followed the same temporal information boundary and post97 evaluation dates as the frozen pathways. Their configurations were determined using the same 60-day offline configuration period, after which the models were refitted using admissible data available before post97 evaluation. This comparison was conducted as an independent post-configuration benchmark to assess the performance of the frozen pathways relative to contemporary forecasting models while preserving the established pathway-configuration results and selection process.

2.5. Candidate Pathway Space

A candidate pathway combines admissible choices of information, target representation, predictor, temporal representation, and output rule. Table 1 first distinguishes the

three dataset-specific weather interfaces and then summarizes the common configuration dimensions.

Table 1. Dataset-specific weather interfaces and configuration dimensions of the candidate pathway space.

Panel A. Dataset-Specific Weather-Interface Definitions			
Interface	Dataset	Exact Definition	Weather Timing
F2-Regional	Internal regional renewable-generation data from China	24 target-day weather-forecast features retrospectively assembled from 13 registered regional weather columns, together with 6 calendar variables.	Archived previous-run target-day forecasts with an approximately one-day nominal lead.
F2-RTE-Hist	External RTE six-target data	13 lagged and rolling historical-weather features joined to the forecast-origin-safe daily RTE matrix.	Past weather information available before the forecast origin.
W2-RTE	External RTE retrospective weather experiments	20 target-day full-weather/ERA5 variables used by C1–C4 capacity-normalized candidates.	Retrospective target-day weather/reanalysis.
Panel B. Common Configuration Dimensions			
Configuration Dimension	Candidate Options	Methodological Role	
Information interface	Lagged target and cross-target histories; calendar variables; plan-derived operating context; F2-Regional archived forecast inputs; F2-RTE-Hist; W2-RTE retrospective inputs; system-state descriptors	Defines admissible explanatory information and its timing boundary	
Target representation	Level; Delta; previous-year-capacity-normalized 24 h daily utilization (CapF) for eligible external generation targets	Determines whether the pathway models absolute output, inter-day change, or normalized utilization	
Predictor family	Persistence; Ridge; RF; XGBoost; LightGBM; representation-enhanced models	Determines the forecasting mechanism	
Representation enhancement	None; Transformer latent vector; PCA-compressed latent vector	Controls temporal representation complexity	
Output rule	Direct output; persistence-based guard	Determines final point-forecast production	
Configuration procedure	Chronological validation; refitting; pathway freezing	Determines and fixes the target-specific pathway before evaluation	

Candidate sets are defined separately for each target. Combinations requiring unavailable or physically irrelevant inputs are excluded before validation, and the remaining pathways are compared under the same chronological protocol.

Model complexity is therefore treated as a configuration choice. The four-target and external candidate sets are specified in the corresponding experimental protocols.

Output rules are target-specific and pre-specified. The hydropower guard is attached only after an eligible Delta pathway has been selected and after the predictor has produced its raw forecast; it is not ranked as a separate predictor, and pathway selection uses the pre-guard prediction errors.

The proposed framework selects a target-specific forecasting pathway from a finite set of admissible candidates through chronological validation. The admissible candidate space is constructed before validation according to target-specific physical relevance and forecast-

origin information availability. Controlled pathway comparisons and candidate-space constraints are used to examine the effects of the information interface, target representation, predictor family, optional temporal representation, training-history length, and candidate-family availability. Under consistent chronological boundaries, these analyses show how constraints on the candidate space affect pathway selection and subsequent forecasting performance, thereby clarifying the role of each configuration element in target-specific forecasting.

2.5.1. Exact Construction of ProvincePlanT+Lag

ProvincePlanT+Lag is represented by a 13-dimensional plan-information vector. The symbols for the target-day province-level plan, calendar month, and previous-day target observation are defined in Supplementary Table S1. For a historical plan window of length w , the following index set identifies the non-missing historical plan observations within the window ending at the previous day:

$$v_{d,w} = \{i \in \{1, \dots, w\} : P_{d-i} \neq \text{NA}\}, w \in \{3, 7, 14, 28\} \quad (12)$$

where $v_{d,w}$ denotes the index set of non-missing historical plan observations within the rolling window; d denotes the target date; w denotes the rolling-window length in days; i denotes the backward lag index; P_{d-i} denotes the province-level plan value recorded i days before target date d ; and NA denotes a missing plan value.

The number of non-missing observations in the window and the corresponding minimum-observation requirements are defined as follows:

$$n_{d,w} = |v_{d,w}|, m_w = \max\{2, \min(3, w)\} \quad (13)$$

where $n_{d,w}$ denotes the number of non-missing plan observations in $v_{d,w}$; $|v_{d,w}|$ denotes the cardinality of this index set; and m_w denotes the minimum number of valid observations required by the implementation for a rolling window of length w .

The rolling plan mean is then computed from the non-missing historical plan values as follows:

$$\mu_{P,d-1}^{(w)} = \frac{1}{n_{d,w}} \sum_{i \in v_{d,w}} P_{d-i}, n_{d,w} \geq m_w \quad (14)$$

where $\mu_{P,d-1}^{(w)}$ denotes the rolling mean of the province-level plan over the w -day historical window ending at $d - 1$. The condition $n_{d,w} \geq m_w$ indicates that the rolling mean is computed only when the minimum-observation requirement is satisfied.

If the number of non-missing observations is below the required minimum, the corresponding rolling feature remains missing. The rolling statistics are calculated after shifting the plan series by one day. Therefore, the target-day plan is retained as a separate plan component and does not enter the historical rolling statistics. The complete ProvincePlanT+Lag vector is written compactly as follows:

$$\mathbf{p}_{j,d} = [f_{1,j,d}, f_{2,j,d}, \dots, f_{13,j,d}]^T \quad (15)$$

where $\mathbf{p}_{j,d}$ denotes the 13-dimensional ProvincePlanT+Lag vector for forecasting target j on target date d ; $f_{k,j,d}$ denotes the k -th registered component; $k = 1, \dots, 13$ indexes the component order; j identifies the forecasting target; and the superscript T denotes transposition, so the 13 components are arranged as a column vector.

The 13 vector components follow the implementation order reported in Supplementary Table S1. The target-day plan components use the target-day province plan available at the

forecast origin, whereas the lagged, rolling, and historical-change terms use historical plan records through the previous day.

2.5.2. Exact Construction and Chronological Generation of PSC

PSC is represented by a 13-dimensional chronological predicted-state vector constructed from four upstream predicted states: thermal generation (TH), hydropower generation (HY), imported electricity (IM), and renewable generation (RE). TH, HY, IM, and RE use the same physical labels as the corresponding internal targets defined in Section 2.1. For each upstream state, the target-day predicted state is generated within the forecast-origin information boundary, and its Delta-type PSC component is formed relative to the corresponding observed state available on day $d - 1$:

$$\delta_{s,d} = \hat{Y}_{s,d} - Y_{s,d-1}, s \in \{\text{TH, HY, IM, RE}\} \quad (16)$$

where $\delta_{s,d}$ denotes the Delta-type PSC component for upstream state s on target date d ; $\hat{Y}_{s,d}$ is the forecast-origin-valid predicted state for date d ; $Y_{s,d-1}$ is the corresponding observed state available on the previous day; d indexes the target date; and s indexes TH, HY, IM, or RE.

The remaining derived components comprise cross-state shares and balance descriptors, defined as follows:

$$r_{\text{HY},d} = \frac{\hat{Y}_{\text{HY},d}}{|\hat{Y}_{\text{TH},d}| + \varepsilon}, r_{\text{IM},d} = \frac{\hat{Y}_{\text{IM},d}}{|\hat{Y}_{\text{TH},d}| + \varepsilon} \quad (17)$$

where $r_{\text{HY},d}$ and $r_{\text{IM},d}$ denote the hydropower and imported-electricity predicted-state shares relative to the magnitude of the predicted thermal-generation state; $\hat{Y}_{\text{HY},d}$ and $\hat{Y}_{\text{IM},d}$ are the corresponding target-day predicted states; $\hat{Y}_{\text{TH},d}$ is the predicted thermal-generation state; and ε is the numerical stabilizer defined below.

$$b_d = \hat{Y}_{\text{TH},d} - \hat{Y}_{\text{HY},d} - \hat{Y}_{\text{IM},d}, q_d = \frac{b_d}{|\hat{Y}_{\text{TH},d}| + \varepsilon} \quad (18)$$

where b_d denotes the empirical balance descriptor formed from the predicted thermal-generation, hydropower, and imported-electricity states on target date d , and q_d denotes the corresponding normalized descriptor.

$$r_{\text{RE},d} = \frac{\hat{Y}_{\text{RE},d}}{|\hat{Y}_{\text{TH},d}| + \varepsilon}, \varepsilon = 10^{-6} \quad (19)$$

where $r_{\text{RE},d}$ denotes the renewable-generation normalized predicted-state descriptor relative to the magnitude of the predicted thermal generation; $\hat{Y}_{\text{RE},d}$ is the predicted renewable generation; and $\varepsilon = 10^{-6}$ is the numerical stabilizer used in the normalized-descriptor denominators. The normalized and balance terms are empirical descriptors of relationships among predicted states and are interpreted as statistical pathway features rather than physical generation shares or system power-balance identities.

The complete PSC vector follows the implementation order reported in Supplementary Table S2 and is written as follows:

$$s_d^{\text{PSC}} = [\delta_{\text{TH},d}, \hat{Y}_{\text{TH},d}, \delta_{\text{HY},d}, \hat{Y}_{\text{HY},d}, \delta_{\text{IM},d}, \hat{Y}_{\text{IM},d}, r_{\text{HY},d}, r_{\text{IM},d}, b_d, q_d, \hat{Y}_{\text{RE},d}, \delta_{\text{RE},d}, r_{\text{RE},d}]^T \quad (20)$$

where s_d^{PSC} denotes the 13-dimensional PSC feature vector for target date d , the component order follows Supplementary Table S2, and the superscript T denotes transpose.

Training-period PSC states are generated from chronological expanding-window out-of-fold upstream predictions. For each random seed, five chronological holdout segments are formed after the minimum training length, as follows:

$$n_{\min} = \min(\max(180, \lfloor n/3 \rfloor), \max(60, n - 1)) \quad (21)$$

where $n_{\min}$ denotes the minimum number of rows required before the expanding-window upstream-state prediction procedure begins; n is the number of rows in the corresponding outer-training set; $\lfloor n/3 \rfloor$ denotes the greatest integer not exceeding $n/3$; and min and max denote the minimum and maximum operators, respectively. The constants 180 and 60 are the registered lower bounds used by the implementation.

The upstream state model uses XGBoost with random seeds 42, 43, and 44, and the seed-specific state predictions are averaged before the PSC variables are assembled. For each outer-test or fixed-execution block, the upstream state model is fitted on the admissible training rows preceding that block and then used to generate the target-date states. Only deterministic calendar terms and lagged or rolling history of the corresponding upstream state are supplied to the state models. Realized target-day generation values, future observations, and raw target-day weather variables remain outside PSC construction. Supplementary Algorithm S1 reports the complete chronological generation procedure.

2.6. Retrospectively Constructed F2-Regional Weather Interface and Renewable RF Ensemble

The F2-Regional interface was retrospectively constructed from archived target-day forecasts retrieved through the Open-Meteo Previous Model Runs API [45]. For each target day D , variables with the `_previous_day1` suffix represent values predicted 24 h before the valid time, corresponding to an approximately one-day nominal lead [45]. The selected weather fields were combined with six calendar variables. Accordingly, the three-year weather series is treated as an archived forecast-based retrospective input rather than a contemporaneously collected operational weather feed.

For each weather variable, the interface combines eight location-level forecast series, one regional mean, and four subregional means, yielding 13 registered regional weather columns. Because the province-wide and subregional means are derived from the same location-level series, these columns form a hierarchical regional aggregation rather than 13 independent locations:

$$\bar{w}_{v,d} = \frac{1}{|S_{v,d}|} \sum_{s \in S_{v,d}} w_{v,d,s} \quad (22)$$

where v indexes the 24 weather variables, d indexes the target day, and s indexes the 13 registered regional weather columns. $w_{v,d,s}$ denotes the forecast value of weather variable v for target day d from column s , while $S_{v,d}$ is the set of column indices with finite values on that date and $|S_{v,d}|$ is the corresponding number of available columns. The aggregated regional weather feature $\bar{w}_{v,d}$ is therefore calculated as the arithmetic mean of the available registered columns. When all 13 columns are present, $|S_{v,d}| = 13$. Under complete two-location subregions, the implied weights of the eight base locations are equal, whereas missing columns change the effective weights.

The resulting regional target-day weather-forecast interface (F2-Regional) contains 24 aggregated weather variables and six calendar variables:

$$\mathbf{x}_d^{F2} = [\hat{\mathbf{w}}_d^T, \mathbf{c}_d^T]^T \in \mathbb{R}^{30} \quad (23)$$

where the weather block contains the 24 hierarchically aggregated fixed-24-h-lead target-day forecast variables, the calendar block contains six variables, and the complete F2-

Regional vector is 30-dimensional. The exact implemented definitions and Open-Meteo source fields of the 24 weather features are listed in Supplementary Table S3.

Three Random Forest models use the same samples, F2-Regional features, and hyper-parameters and differ only in seed. Their predictions are combined by arithmetic averaging:

$$\hat{y}_d^{RE} = \frac{1}{R} \sum_{r=1}^R f_{\text{RF}}^{(r)}(\mathbf{x}_d^{F2}), R = 3 \quad (24)$$

where $\hat{y}_d^{RE}$ is the final renewable-generation forecast for date d ; $f_{\text{RF}}^{(r)}(\cdot)$ is the Random Forest predictor trained with the r -th random seed; r is the random-seed index; R is the number of repeated Random Forest models; and $\mathbf{x}_d^{F2}$ is the forecast-origin feature vector defined in Equation (23). Here, $R = 3$, corresponding to random seeds 42, 43, and 44.

2.7. Pathway Selection and Configuration Freezing

For target j , each candidate combines an information set, target representation, predictor family, optional latent representation, and output rule:

$$p_j = (I_j, R_j, M_j, H_j, O_j) \in P_j \quad (25)$$

where p_j denotes one candidate pathway for target j ; I_j is the admissible forecast-origin information set; R_j is the target representation, such as Level or Delta; M_j is the predictor family; H_j is the optional latent-representation component; O_j is the output rule; and P_j is the target-specific candidate pathway space.

Before validation, pathways are screened for physical relevance and information availability at the forecast origin. The four-target study uses 4 April–2 June 2025 ($N = 60$) as the formal offline configuration period. The later fixed95 interval is used only for common-date developmental comparison, whereas the post97 period remains fully subsequent and is excluded from pathway selection. The complete evidence hierarchy is summarized in Table 2.

$$\mathcal{A}_j = 1 \{ \widehat{\Pr}(\Delta_j < 0) \geq 0.95 \wedge U_{0.95}\Delta_j < 0 \}, \Delta_j = \text{RMSE}_j^{\text{alt}} - \text{RMSE}_j^{\text{ret}} \quad (26)$$

Table 2. Evidence hierarchy and chronological role of the internal four-target data stages.

Stage	Date Range (N)	Training/Configuration Boundary	Analytical Role	Selection Use and Overlap
60-day offline configuration	4 April–2 June 2025 (60)	Historical data available before 4 April 2025	Formal offline pathway configuration	Yes; sole formal internal selection window
fixed95	3 June–5 September 2025 (95)	Frozen paths refitted before 3 June 2025	Common-window developmental comparison and reporting	No; overlaps T4 and T5
T3	1 January–31 March 2025 (90)	Training data through 31 December 2024	Pre-configuration retrospective diagnostic	No; precedes 60-day configuration
T4	1 April–30 June 2025 (91)	Training data through 31 March 2025	Configuration-overlap retrospective diagnostic	No; overlaps 60-day and fixed95
T5	1 July–5 September 2025 (67)	Training data through 30 June 2025	Later retrospective diagnostic	No; contained within fixed95
post97	6 September–11 December 2025 (97)	Frozen paths refitted through 5 September 2025	Fully subsequent frozen-path confirmation	No; fully subsequent to all developmental stages

Let Δ_j denote the paired-bootstrap RMSE difference between an alternative pathway and the retained pathway for target j . The alternative is accepted only when the indicator in Equation (26) equals one.

For the four-target study, Equation (26) was applied together with the target-specific evidence specified before the subsequent 97-day evaluation: factor and seed/window evidence for renewable generation, latent-representation evidence for hydropower, and direct candidate comparisons for thermal generation and imported electricity.

For each external target and rolling-origin fold, admissible candidates are ranked by inner-validation RMSE. Complexity rank and candidate identifier are used only when two candidates have identical inner errors. Historical-weather candidates and retrospective target-day weather candidates are evaluated in their respective information settings.

$$\hat{y}_{j,d} = F_j^*(x_{j,d}), d \in E_j \quad (27)$$

where $\hat{y}_{j,d}$ is the forecast for target j on date d ; $F_j^*(\cdot)$ is the forecasting function associated with the pathway selected and frozen for target j ; $x_{j,d}$ is the forecast-origin input vector for target j on date d ; and E_j is the designated evaluation-date set for target j .

After selection, the information interface, target representation, preprocessing, model family, hyperparameters, seed treatment, latent dimension, and output rule are fixed. Model parameters are refitted using admissible pre-evaluation data, and the frozen pathway is applied to the designated evaluation dates. The annual external sequence is repeated independently in each rolling-origin fold; the 2022-selected paths are refitted before 2023 and held fixed throughout the 2023 confirmation.

2.8. Chronological Configuration and Evaluation Protocol

2.8.1. Four-Target Chronological Protocol

The calendar span from 1 January 2023 to 5 September 2025 contains 979 dates. The source archive includes 948 recorded dates, with a 31-day gap from 1 to 31 May 2023. Alignment across the four finite target series yields 945 common daily observations; three dates at the end of December 2024 lack complete observations across the regional targets. Target-specific modeling uses the admissible observations available for each target. A separately locked confirmation archive contains the following 97 days, from 6 September to 11 December 2025, and provides the fully subsequent evaluation. All fitting, preprocessing, pathway comparison, and refitting follow chronological order, and learned transformations are estimated from data available before the corresponding stage.

The internal evidence hierarchy contains one formal 60-day offline pathway-configuration period, the fixed95 common-window developmental comparison and reporting period, three retrospective diagnostic windows (T3–T5), and the post97 fully subsequent frozen-path confirmation. T4 overlaps the 60-day configuration and fixed95 periods, whereas T5 is contained within fixed95. Fixed95, T3, T4, and T5 provide complementary developmental or diagnostic evidence and do not replace the 60-day selection layer. The subsequent T3–T5 periods are used for retrospective temporal diagnostics to evaluate the stability and temporal generalization characteristics of the configured pathways across different time stages.

Pooled T3–T5 metrics are recomputed from concatenated daily observations and forecasts and summarize the combined retrospective diagnostic record. They are reported separately from the fixed95 developmental comparison and the fully subsequent post97 confirmation.

The internal chronological configuration and evaluation protocol is summarized in Figure 2a.

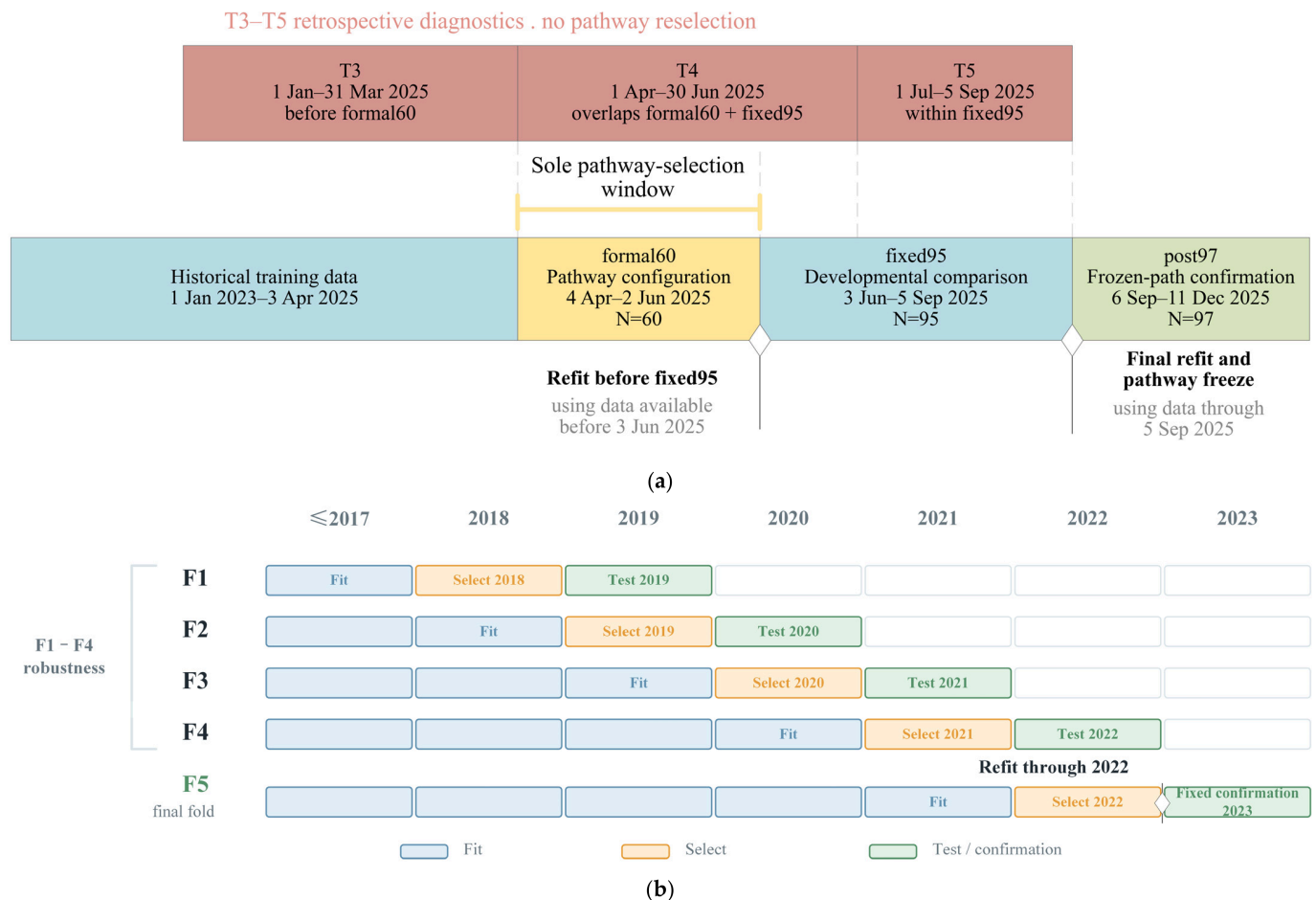


Figure 2. Unified chronology of the internal and external configuration and evaluation protocols. (a) Internal four-target study: the timeline includes the T3–T5 retrospective diagnostic windows, the formal60 pathway-selection window, the fixed95 developmental comparison, the refitting and freezing boundaries, and the fully subsequent post97 confirmation. T4 overlaps formal60 and fixed95, whereas T5 is contained within fixed95. (b) External RTE study: F1–F4 provide rolling robustness evidence through annual fit-select-test progression, whereas F5 uses 2022 as the final selection year, followed by refitting through 2022 and fixed-path confirmation in 2023.

Before the formal60 pathway-configuration window illustrated in Figure 2a, the effective training sets contained 792, 790, and 782 samples for the thermal, renewable, and hydropower candidate models, respectively. For imported electricity, the Level and Delta candidate pathways used 790 and 788 effective training samples, respectively, while the Persistence pathway generated forecasts directly from the previous-day observation. The formal60 window then provided 60 comparison dates for each target. After pathway selection, the retained thermal, renewable, and hydropower models were refitted using all admissible data available through 5 September 2025, corresponding to 947, 945, and 937 effective refitting samples, respectively. The retained imported-electricity Persistence pathway continued to generate forecasts directly from the $D - 1$ observation; before post97, 945 historical observations were available, with 943 satisfying the $D - 1$ anchoring requirement.

After the effective samples were determined under the $D - 1$ information boundary, target variables and input features were processed according to their respective data characteristics. Target labels were not imputed; records with non-finite target values or that were missing the required $D - 1$ anchor observations were excluded from supervised training and model evaluation. No explicit outlier deletion was applied to the internal

target time series. For numerical features used by tree-based models, missing values were handled using training-set-only median statistics within the fitted preprocessing pipeline. For the F2-Regional weather interface, daily regional weather indicators were calculated from available finite registered weather columns. In the external RTE dataset, physically implausible relative-humidity values were clipped to the range of 0–100%, and small negative generation measurements were corrected to zero during target cleaning.

Under the same final-refit boundary described above, runtime was evaluated for the retained Transformer-based hydropower pathway, PCA16-TXGB-Delta + ProvincePlanT+Lag + output guard. The tests were performed on an Intel Core i7-14700KF workstation with 63.84 GB of RAM, and the Transformer, PCA, and XGBoost stages all ran on the CPU. Averaged over three runs, Transformer fitting and latent-feature generation took 1.457 ± 0.013 s, followed by 0.0066 ± 0.0002 s for PCA fitting and transformation and 0.239 ± 0.017 s for XGBoost fitting. The full refitting procedure took 1.705 ± 0.021 s. Forecast generation for all 97 post97 days took 0.0299 ± 0.0008 s in total, equivalent to 0.000308 ± 0.000008 s per forecasting day.

2.8.2. External RTE Protocol

The external analysis uses the national consolidated and definitive éCO2mix dataset released by the French transmission system operator [46]. Half-hourly timestamps are interpreted in Coordinated Universal Time and converted to the Europe/Paris local calendar before daily aggregation. Valid power records are converted from MW to MWh using a 0.5 h factor and summed by local day. Daily source-record counts and weather-coverage statistics are retained for quality control, and local daylight-saving days with 23 or 25 accumulated hours remain in the dataset.

Across 2012–2023, the dataset contained 24 daylight-saving transition days. As a representation-level sensitivity check, CapF was additionally recalculated on these dates using the actual local-day duration of 23 or 25 h in place of the fixed 24 h denominator. For the affected renewable, solar, and wind records on these transition dates, the maximum relative difference from the fixed-24 h representation was approximately 4.35%, and the mean absolute relative difference was approximately 4.17%.

Fossil thermal generation is defined as the sum of oil-, coal-, and gas-fired generation; renewable generation is the sum of wind and solar generation; hydropower, solar generation, and wind generation are also evaluated separately. The signed physical-exchange series is retained as net physical exchange so that positive and negative values preserve flow direction. The resulting daily dataset contains 4383 dates from 1 January 2012 to 31 December 2023. Five outer folds evaluate 2019–2023 and provide 1826 aligned daily predictions for each target.

Annual installed capacities were obtained from the RTE/ODRÉ national annual generation-capacity dataset, which reports sector capacity at 31 December [47]. For predictions in year y , CapF uses the registered capacity at the end of $y - 1$ and a fixed 24 h denominator; renewable capacity is the sum of wind and solar capacity. Both external weather interfaces were derived from ERA5. F2-RTE-Hist contains 13 lagged and rolling historical-weather variables constructed only from weather information available through $D - 1$, whereas W2-RTE contains 20 retrospective target-day ERA5 variables and is used only by C1–C4 in the separately reported retrospective complete-weather benchmark. Generic CapF candidates do not automatically include W2-RTE, and net physical exchange is ineligible for CapF. The exact weather-variable names and daily aggregation rules are listed in Supplementary Table S4.

The external protocol comprises five RTE rolling-origin outer test folds. F1 selects on 2018 and tests 2019; F2–F5 extend the same sequence by one year and test 2020–2023. All

preprocessing and fitting procedures use only the corresponding pre-test subset. F1–F4 provide rolling robustness evidence, while F5 uses 2022 inner evidence, refits before 2023, and keeps the selected path fixed throughout the held-out 2023 evaluation. For W2-RTE, the associated target-day ERA5 covariates are supplied retrospectively.

The corresponding external rolling-origin chronology is summarized in Figure 2b.

2.8.3. Controlled Training-History Reduction

The internal sample-depth experiment refits the frozen pathways with the full admissible history and with the most recent 450, 365, or 270 training days, while keeping the post97 evaluation dates unchanged. In the external main analysis, the retrospective target-day weather pathways are refitted with the full history and with the most recent 365 days. Forecast-origin-safe training-history results are retained as supplementary boundary evidence.

2.8.4. Controlled Input-Interface Incompleteness

The main information-unavailability experiment removes the capacity-normalized candidate family before fold-specific selection and then repeats configuration and refitting. For renewable, solar, and wind generation, this removal excludes the retrospective target-day weather candidates because they use the capacity-normalized representation. A separate removal of the historical-weather interface is reported in the Supplementary Table S5. The reported changes therefore describe reconfiguration after candidate eligibility changes, rather than the isolated effect of a single input variable.

2.8.5. Decoupled Factorial Ablation

To quantify the individual contributions of the main pathway components, a decoupled factorial ablation was conducted for the external renewable, solar, and wind targets. Four factors were crossed independently: capacity scaling (Raw or CapF), retrospective target-day weather availability (no-W2 or W2-RTE), target representation (Level or Delta), and predictor family (XGBoost or TXGB). Here, Raw/CapF specifies whether the target is modeled on its original scale or after previous-year-capacity normalization, whereas Level/Delta specifies whether the modeled quantity is represented as its absolute value or inter-day change. The no-W2 setting zero-masks the 20 retrospective target-day W2-RTE variables while retaining the remaining input structure unchanged, whereas the W2-RTE setting supplies these variables retrospectively. The resulting $2 \times 2 \times 2 \times 2$ design produced 16 predefined configurations for each target and fold. All configurations were evaluated over the same five rolling-origin outer folds, yielding 240 target-fold fits in total.

For each factor, matched comparisons were constructed by changing only that factor while keeping the other three factors unchanged. Capacity-normalization effects were estimated from Raw–CapF pairs with identical weather, representation, and predictor settings. Raw and CapF pairs retained the same exogenous information and predictor setting, with the contrast changing only the target scaling and its corresponding reconstruction. Weather effects were estimated from no-W2–W2-RTE pairs with identical capacity, representation, and predictor settings; representation effects from Level–Delta pairs with identical capacity, weather, and predictor settings; and predictor-family effects from XGBoost–TXGB pairs with identical capacity, weather, and representation settings. The factorial analysis was conducted as a post-configuration diagnostic using the predefined model settings. No additional pathway selection was performed, and the resulting evidence was retained separately from the original pathway-selection process.

Factor effects were summarized using matched out-of-sample RMSE differences. A 7-day moving-block bootstrap with 5000 repetitions was used to account for temporal dependence and to quantify uncertainty in each matched effect. W2-RTE represents retro-

spective target-day ERA5 reanalysis information and serves as an information-availability benchmark in this analysis.

2.8.6. Repeated-Window Pathway-Selection Stability

To evaluate the stability of the 60-day internal configuration decision under repeated short-window comparisons, a repeated-window analysis was conducted for all four internal targets. Eight chronological 60-day configuration windows were registered for each target, with successive windows shifted by approximately 30 eligible observations. All primary windows ended on or before 2 June 2025, so fixed95 and post97 data remained outside the stability analysis. Candidate eligibility, preprocessing, and the registered target-specific decision rule were reconstructed within each window using only training information available before that window.

For each frozen-pathway versus alternative comparison, paired RMSE differences were evaluated using a 7-day moving-block bootstrap with 5000 repetitions to preserve short-range temporal dependence. Within each target-window family, the resulting one-sided bootstrap p -values were adjusted by Holm's step-down procedure at $\alpha = 0.05$ to control the family-wise error rate across multiple simultaneous candidate comparisons. Stability was summarized by the frequency with which the reconstructed decision rule retained the original formal60 pathway, the frozen-path rank and regret relative to the point-RMSE winner, and the number of windows containing any Holm-corrected alternative replacement evidence. A four-window non-overlapping subset was additionally examined as a sensitivity check. This analysis served as post-configuration revision evidence and did not alter the original formal60 selections or any subsequent frozen-path evaluation.

2.9. Evaluation, Interpretation, and Uncertainty Quantification

2.9.1. Point-Forecast Metrics

Point forecasts are evaluated using RMSE as the principal error measure, with MAE and R^2 used as complementary measures of average absolute deviation and stage-relative explanatory performance, respectively:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (28)$$

where N is the number of evaluated dates; y_i is the observed target value for date i ; and $\hat{y}_i$ is the corresponding point forecast.

To facilitate comparison across targets with different output scales, normalized RMSE (nRMSE) is additionally reported for each target. The normalization uses the mean absolute observed value over the corresponding evaluation record.

$$\text{nRMSE}(\%) = 100 \times \frac{\text{RMSE}}{\frac{1}{N} \sum_{i=1}^N |y_i|} \quad (29)$$

where N is the number of evaluated dates and y_i is the observed target value for date i ; the mean absolute observed value is used as the denominator so that the normalized metric remains applicable to targets that may contain both positive and negative values.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (30)$$

where N , y_i , and $\hat{y}_i$ have the same meanings as in Equation (28). MAE measures the average absolute magnitude of the forecast errors.

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (31)$$

where $\bar{y}$ is the mean of the observed values in the corresponding evaluation stage; the other symbols are defined above. R^2 measures performance relative to the evaluation-stage mean benchmark.

All metrics are computed separately for each target and evaluation stage. Bias is the mean of prediction minus observation. RMSE skill relative to Persistence is calculated as $100 \times \frac{\text{RMSE}_{\text{Persistence}} - \text{RMSE}_{\text{path}}}{\text{RMSE}_{\text{Persistence}}}$. Key paired comparisons use 5000 bootstrap repetitions, with fold-stratified resampling for the external rolling evaluation.

2.9.2. Target-Appropriate Interpretation

TreeSHAP is applied to the no-latent Plan-XGBoost-Delta thermal model and the pre-guard XGBoost core of the hydropower TXGB model. Renewable generation is interpreted through controlled comparisons of the F2 and PSC information paths and the main effects of information, model, and target representation. Imported electricity is examined through prediction errors and day-to-day changes.

2.9.3. Chronological Out-of-Sample Residual Intervals

Empirical chronological out-of-sample absolute-residual intervals are constructed around the frozen point forecasts. Calibration uses only out-of-sample residuals generated before the corresponding evaluation period. The renewable calibration data contain 533 residual records spanning 357 dates; when several inner blocks contribute residuals for the same date, they are averaged to form one date-balanced calibration residual. Hydropower calibration uses 95 out-of-sample residuals from the fixed 95 dates and applies the resulting interval only to the subsequent post97 period.

$$a_i = |y_i - \hat{y}_i| \quad (32)$$

where i indexes a registered chronological out-of-sample calibration record, y_i is the observed value, and $\hat{y}_i$ is the corresponding out-of-sample forecast.

Let the absolute residuals be ordered increasingly. For nominal miscoverage α , the bounded one-based rank and interval half-width are as follows:

$$k = \min\{[(n_{\text{res}} + 1)(1 - \alpha)], n_{\text{res}}\}, \hat{q} = a_{(k)} \quad (33)$$

where n_{res} is the number of calibration residuals, k is the bounded one-based order-statistic rank, and $\hat{q}$ is the k -th ordered absolute residual. The unrounded $\hat{q}$ is used in all coverage and width calculations.

For evaluation date d , the empirical interval is centered on the frozen point forecast with lower and upper bounds $\hat{y}_d - \hat{q}$ and $\hat{y}_d + \hat{q}$, respectively:

$$\mathcal{I}_d = [\hat{y}_d - \hat{q}, \hat{y}_d + \hat{q}] \quad (34)$$

where $\mathcal{I}_d$ is the prediction interval for evaluation date d , $\hat{y}_d$ is the frozen point forecast, and $\hat{q}$ is the residual quantile defined in Equation (33).

Empirical coverage and interval sharpness are summarized by PICP and MPIW:

$$\text{PICP} = \frac{1}{N} \sum_{d=1}^N 1\{y_d \in \mathcal{I}_d\} \quad (35)$$

where $1\{\cdot\}$ is the indicator function, $\mathcal{I}_d$ is the prediction interval for evaluation date d , and N is the number of evaluated dates.

$$\text{MPIW} = \frac{1}{N} \sum_{d=1}^N (U_d - L_d) \quad (36)$$

where L_d and U_d are the lower and upper bounds of the prediction interval for evaluation date d , and N is the number of evaluated dates.

Within this subsection, residual-record index i denotes calibration rows, whereas date index d denotes evaluation days. PICP and MPIW summarize empirical chronological coverage and mean interval width; no formal coverage guarantee is assigned to these target-specific residual sources.

To examine whether a fixed residual width remains adequate as forecast errors evolve over time, the uncertainty analysis was extended to two sequential 60-day calibration schemes applied to the unchanged frozen point forecasts: rolling conformal calibration (RC60) and adaptive conformal calibration (ACI60). FIXED uses the pre-post97 calibration set without sequential updating, whereas RC60 recalculates the finite-sample residual quantile from the most recent 60 admissible residual dates before each forecast date. ACI60 uses the same 60-date rolling residual pool and further updates the effective miscoverage level after each newly observed outcome.

For ACI60, the nominal miscoverage level is $\alpha = 0.10$, with an update step of $\gamma = 0.01$. After each observation becomes available, the effective level for the next forecast date is updated by adding $\gamma(\alpha - m_d)$ to the current value and restricting the result to $[0.01, 0.50]$, where $m_d = 1$ when the realized observation falls outside the interval and $m_d = 0$ otherwise. Each interval is determined from prior calibration information, with the observation on date d incorporated into subsequent updates.

Interval performance is evaluated using prediction interval coverage probability (PICP), mean prediction interval width (MPIW), and mean interval score (MIS). MIS summarizes the coverage–sharpness trade-off by combining interval width with penalties for observations falling below the lower bound or above the upper bound. Lower MIS values indicate more favorable overall interval performance.

Additional robustness checks examine directional asymmetry and forecast-magnitude dependence. The asymmetric rolling and adaptive variants estimate the lower and upper signed-residual tails separately. For the magnitude-normalized variants, the calibration score for target date d is defined as follows:

$$r_d = \frac{|e_d|}{\max(|\hat{y}_d|, s_0)} \quad (37)$$

where $e_d = y_d - \hat{y}_d$ is the forecast error, $\hat{y}_d$ is the frozen point forecast, and s_0 is the 5th percentile of the absolute date-balanced calibration forecasts. All robustness variants retain the same frozen point forecasts and use only residual information available before each post97 forecast date. Detailed results are reported in Supplementary Tables S6 and S7.

A weather-conditioned robustness diagnostic was additionally conducted for renewable generation using two forecast-origin-valid F2-Regional variables: daily mean 100 m wind speed and daily shortwave-radiation sum. For each variable, the median over the 357 pre-post97 date-balanced calibration dates defined a fixed Low/High threshold, with Low values below the median and High values at or above it. The existing FIXED, RC60, and ACI60 intervals were then evaluated within each regime while retaining their original frozen point forecasts and interval bounds. This diagnostic therefore characterizes conditional interval behavior under different meteorological states while preserving the

chronological calibration protocol. The regime definitions, sample composition, and interpretation are reported in Supplementary Table S8.

2.10. Implementation Settings

All records are ordered chronologically, and missing-value handling, scaling, Transformer learning, and PCA fitting are estimated from training data only. Table 3 reports the core implementation settings, information boundaries, and pre-specified controls. Detailed model-family and hyperparameter specifications are provided in Supplementary Table S9, while target-specific candidate eligibility is summarized in Supplementary Table S10.

Table 3. Core implementation settings, interface boundaries, and pre-specified controls.

Component	Core Setting
Forecast horizon	One day ahead
Transformer candidate input length	7 days
Transformer candidate architecture	Model dimension 32; 2 attention heads; 1 encoder layer; feed-forward dimension 64
Transformer regularization	Dropout 0.15; GELU activation
Hydropower latent compression	16 PCA components
Renewable Random Forest	300 trees; maximum depth 6; minimum samples per leaf 5
Renewable RF seeds	42, 43, and 44; arithmetic prediction average
XGBoost learning structure	Learning rate 0.04; maximum depth 3; minimum child weight 3
XGBoost sampling	Subsample 0.85; column subsample 0.85
XGBoost regularization	L1 coefficient 0.1; L2 coefficient 1.0
Imported-electricity predictor	Persistence; no fitted model hyperparameters
External CapF denominator	Previous-year forecast-origin-safe registered capacity $\times$ fixed 24 h; reported as previous-year-capacity-normalized 24 h daily utilization.
Hydropower output guard	Replace a negative or non-finite reconstructed forecast with the $D - 1$ Persistence value
External retrospective target-day weather candidates	C1 W2-RTE-CapF-XGBoost-Level; C2 W2-RTE-CapF-XGBoost-Delta; C3 W2-RTE-CapF-TXGB-Level; C4 W2-RTE-CapF-TXGB-Delta
Training-history settings	Four-target FULL/N450/N365/N270; external FULL/N365 with separate fold-specific nested selections and unchanged outer dates
Paired uncertainty analysis	5000 bootstrap repetitions; fold-stratified resampling for external comparisons
External annual fold boundaries	F1: fit $\leq$ 2017, select 2018, refit $\leq$ 2018, test 2019; F2: fit $\leq$ 2018, select 2019, refit $\leq$ 2019, test 2020; F3–F5 extend this sequence through the 2022 $\rightarrow$ 2023 fold

The implementation settings of TimeXer and PatchTST used in the post97 contemporary-baseline comparison, together with the other reproducibility settings, are summarized in Supplementary Table S9.

3. Results

3.1. Chronological Pathway Selection for the Four Targets

All admissible pathways were evaluated on the same 60-day offline configuration period. RMSE provided the common numerical comparison, and close alternatives were assessed with the paired-bootstrap criterion defined in Section 2.7. When multiple pathways

showed comparable performance, pathway selection was determined by considering both point forecasting performance and the statistical support of performance differences. Table 4 reports the retained pathway, the nearest admissible alternative, and the paired uncertainty for each target. The fixed95 and post97 periods were excluded from this selection.

Table 4. Target-specific pathway decisions from the 60-day offline configuration period.

Target	Retained Pathway	60-Day RMSE (MWh)	Nearest Alternative (RMSE, MWh)	Alternative – Retained Δ RMSE [95% Interval] (MWh)
Local thermal generation	Plan-XGB-Delta	157.634	LightGBM-Native-Delta (180.353)	+22.719 [9.880, 33.072]
Renewable generation	F2-Regional-RF-Level (seeds 42/43/44)	3191.163	F2-Regional-XGBoost-Level (3132.722)	−58.441 [−245.148, 137.051]
Hydropower generation	PCA16-TXGB-Delta + output guard	256.942	RF-Native-Delta (259.501)	+2.558 [−24.081, 25.638]
Imported electricity	Persistence	2285.279	Ridge-Native-Delta (2316.256)	+30.977 [−98.647, 220.724]

Note: Δ RMSE and its paired interval are defined as alternative minus retained, consistent with Equation (26). Negative values indicate lower RMSE for the alternative pathway. The fixed95 and post97 periods were excluded from pathway selection, and the hydropower guard was applied only after model output was produced.

For renewable generation, F2-Regional-XGBoost-Level achieved a 58.441 MWh lower point RMSE than F2-Regional-RF-Level during formal60. However, the paired-bootstrap interval reported in Table 4 included zero, and the estimated improvement probability of 0.7336 remained below the pre-specified 0.95 replacement threshold. Equation (26) was therefore not satisfied, and F2-Regional-RF-Level was retained. Hydropower retained PCA16-TXGB-Delta, with the guard applied only as an output rule. Imported electricity retained Persistence under the available information.

The target-specific pathway decisions above were established using the formal60 configuration window. To assess their robustness to the temporal placement of the 60-day configuration period, the same registered decision logic was reapplied across eight rolling 60-day windows for each internal target. Selection stability was evaluated using pathway-retention frequency together with dependence-aware and multiplicity-controlled alternative comparisons.

As summarized in Table 5, the original formal60 pathway was retained in all eight thermal windows, six of eight renewable windows, six of eight hydropower windows, and seven of eight imported-electricity windows. The candidate with the lowest point RMSE changed in some short windows, indicating moderate sensitivity to window placement. After the 7-day moving-block bootstrap and Holm correction, alternative replacement evidence appeared in none of the thermal or renewable windows and in one hydropower and one imported-electricity window. The four-window non-overlapping subset produced no Holm-corrected replacement evidence for any target.

As a separate sensitivity analysis, a dependence-aware 7-day moving-block bootstrap was applied to the same 60 formal60 dates. The resulting RMSE-difference interval was [−268.668, 135.807] MWh and included zero. A one-sided test based on the centered squared-error differential yielded an unadjusted *p*-value of 0.2773, which became 1.0000 after Holm correction across the 21 pre-specified candidate comparisons. These results provided no multiplicity-controlled evidence for replacing the retained RF pathway.

The four-window non-overlapping sensitivity analysis likewise identified no alternative pathway that remained significant after Holm correction for any target.

Table 5. Repeated-window stability of the internal pathway-selection decisions.

Target	Rolling 60-Day Windows	Original Formal60 Pathway Retained	Windows with Holm-Corrected Alternative Evidence
Thermal generation	8	8/8 (100%)	0/8
Renewable generation	8	6/8 (75%)	0/8
Hydropower generation	8	6/8 (75%)	1/8
Imported electricity	8	7/8 (87.5%)	1/8

Note: Eight chronological 60-day windows were evaluated for each target, and all windows ended on or before 2 June 2025. “Original formal60 pathway retained” denotes the frequency with which the registered decision rule retained the pathway selected in the formal 60-day configuration period. Alternative replacement evidence was assessed using a 7-day moving-block bootstrap with 5000 repetitions followed by Holm correction at $\alpha = 0.05$. A four-window non-overlapping subset was additionally examined as a sensitivity check.

After the 60-day offline configuration, the frozen pathways were refitted using data available before 3 June 2025 and evaluated on the common 95 dates from 3 June to 5 September 2025. This fixed95 period is a developmental comparison and reporting layer. Table 6 summarizes the archived daily-prediction metrics and remains separate from post97 confirmation.

Table 6. Performance of the retained pathways during the aligned 95-day developmental comparison window.

Target	Frozen Pathway	RMSE (MWh)	MAE (MWh)	R ²	Bias (MWh)	RMSE Skill
Local thermal generation	Plan-XGB-Delta	188.094	141.962	0.7196	−14.381	13.37%
Renewable generation	F2-Regional-RF-Level	2822.421	2215.023	0.6751	−285.484	40.70%
Hydropower generation	PCA16-TXGB-Delta + guard	350.623	231.302	0.8662	7.962	1.67%
Imported electricity	Persistence	1675.615	1078.947	0.6438	7.368	0.00%

Note: All targets use the same 95 dates. Bias is prediction minus observation. RMSE skill is calculated relative to the target-specific Persistence forecast. The window supports aligned developmental comparison and is separate from the fully subsequent post97 confirmation.

The aligned comparison retained the target-dependent pattern observed during configuration. Renewable generation achieved the largest improvement over Persistence (40.70%), local thermal generation improved by 13.37%, and hydropower provided a smaller 1.67% gain. Imported electricity remained assigned to Persistence. Hydropower achieved the highest R² (0.8662), while the lower renewable R² reflects the greater daily variability of aggregated wind and solar output during this window.

3.2. Chronological Robustness Across T3–T5

Figure 3 presents retrospective stage-wise performance across T3–T5, and Table 7 reports pooled metrics recomputed from concatenated daily observations and forecasts. T3 describes the pre-configuration stage, T4 describes the configuration-overlap stage, and T5 extends the analysis to a later period. Thermal generation remained strong in T3 and T4 but weakened in T5; renewable generation improved after T3; hydropower retained high explanatory performance; and imported electricity varied most across stages.

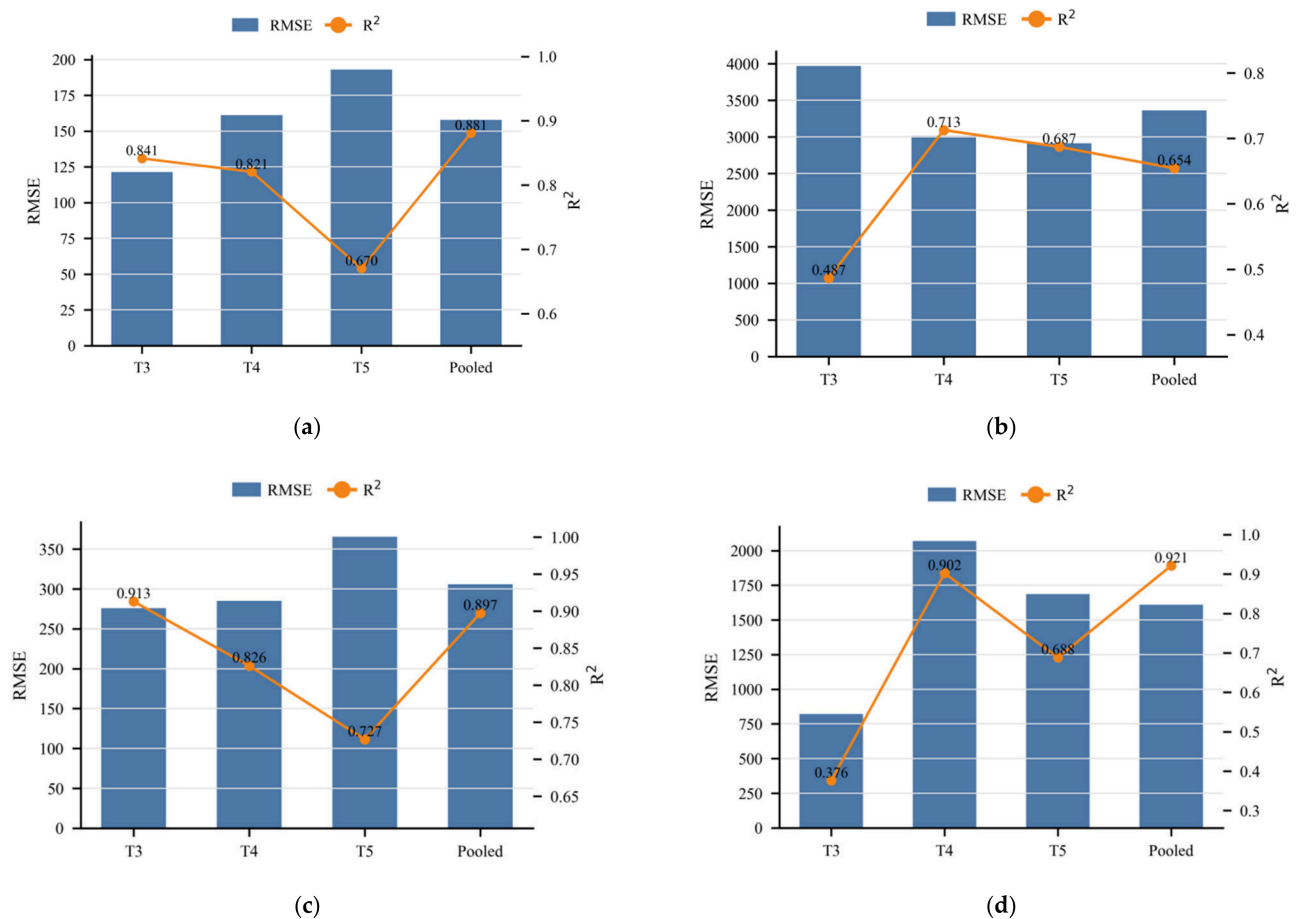


Figure 3. Stage-wise forecasting performance of the four frozen pathways across T3–T5 and the pooled period: (a) local thermal generation; (b) renewable generation; (c) hydropower generation; and (d) imported electricity.

Table 7. Pooled T3–T5 performance and relative RMSE improvement over Persistence.

Target	Frozen Pathway	N	Pooled RMSE (MWh)	Pooled R^2	Persistence RMSE (MWh)	RMSE Skill
Local thermal generation	Plan-XGB-Delta	248	157.883	0.8810	199.397	20.82%
Renewable generation	F2-Regional-RF-Level	247	3364.193	0.6543	5674.307	40.71%
Hydropower generation	PCA16-TXGB-Delta + guard	247	305.927	0.8972	311.573	1.81%
Imported electricity	Persistence	247	1610.014	0.9212	1610.014	0.00%

Note: N denotes aligned daily observations. Pooled T3–T5 metrics summarize the concatenated retrospective diagnostic record. The pooled R^2 reflects both within-stage forecasting performance and variation among the three stages.

Across the concatenated T3–T5 record, the retained pathways achieved RMSEs of 157.883, 3364.193, 305.927, and 1610.014 MWh for local thermal generation, renewable generation, hydropower, and imported electricity, respectively. Figure 3 further shows target-dependent temporal behavior: hydropower remained comparatively stable, renewable forecasting improved after T3, and imported electricity showed the greatest stage-to-stage variation.

3.3. Fully Subsequent 97-Day Confirmation

3.3.1. Point-Forecast Performance

Each selected configuration was refitted using all admissible data available through 5 September 2025 and then evaluated on the next 97 days, from 6 September to 11 December 2025. Table 8 reports the point-forecast metrics, and Figure 4 compares the observed and predicted daily electricity volumes.

Table 8. Performance of the four frozen pathways during the subsequent 97-day confirmation period.

Target	Frozen Pathway	N	RMSE (MWh)	nRMSE (%)	MAE (MWh)	R ²	Bias (MWh)
Local thermal generation	Plan-XGB-Delta	97	148.645	16.02	100.396	0.8084	50.111
Renewable generation	F2-Regional-RF-Level	97	2759.543	21.88	2193.202	0.7337	92.499
Hydropower generation	PCA16-TXGB-Delta + guard	97	429.113	9.82	298.332	0.7782	−62.024
Imported electricity	Persistence	97	1728.416	9.36	1265.979	0.5212	47.423

Note: N denotes the number of daily observations in the confirmation period. All four targets are evaluated on the same 97 dates. nRMSE is calculated as RMSE divided by the mean absolute observed value over the corresponding evaluation record and expressed as a percentage.

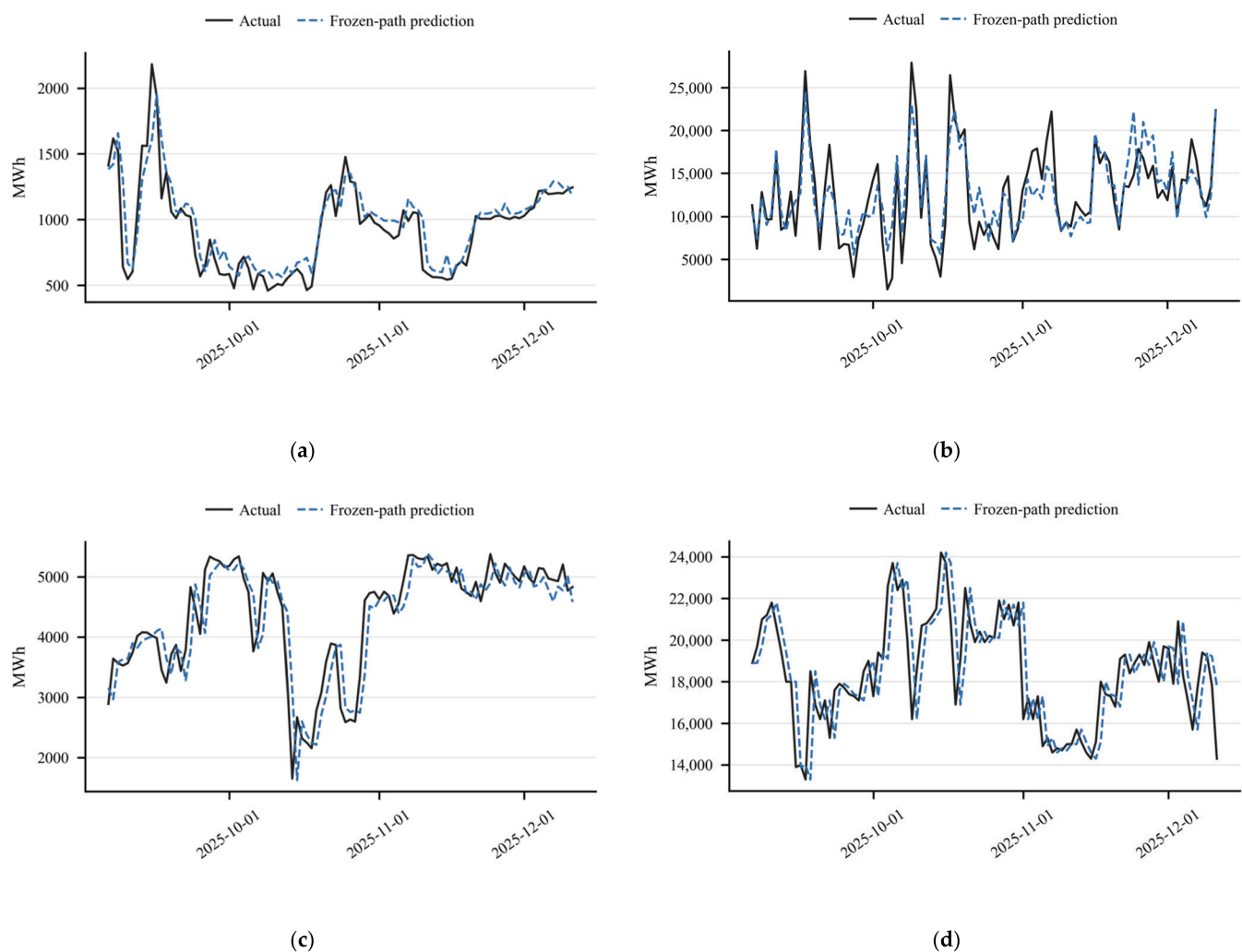


Figure 4. Observed and frozen-path predicted electricity volumes during the subsequent 97-day confirmation period: (a) local thermal generation; (b) renewable generation; (c) hydropower generation; and (d) imported electricity.

The fully subsequent period confirmed strong thermal and renewable performance, with R^2 values of 0.8084 and 0.7337, respectively. Hydropower retained $R^2 = 0.7782$, although its incremental advantage over Persistence remained limited. Imported electricity was less stable in this later period ($R^2 = 0.5212$), consistent with the absence of transaction schedules, tie-line plans, and external dispatch states. In market-oriented use, the thermal and regional supply forecasts provide quantitative inputs for day-ahead volume assessment, whereas the larger exchange errors indicate greater uncertainty in interregional procurement requirements.

3.3.2. Post Hoc Comparison with Pre-Specified Branches and Contemporary Forecasting Baselines on the Subsequent 97-Day Period

After pathway selection and freezing, the pre-specified candidates were further compared over the same 97 subsequent dates to characterize their performance differences. As shown in Figure 5, the retained thermal and renewable pathways reduced RMSE relative to Persistence by 23.2% and 50.9%, respectively. For hydropower, the guard was inactive, so the pre-guard and final forecasts were identical. The no-latent XGB-Delta branch was 1.8% worse than Persistence, whereas PCA16-TXGB-Delta was 0.6% better; the direct comparison corresponds to a 2.44% lower RMSE for the latent pathway. For imported electricity, PCA8-TXGB-Delta produced a 1.2% point improvement, but its paired interval included zero.

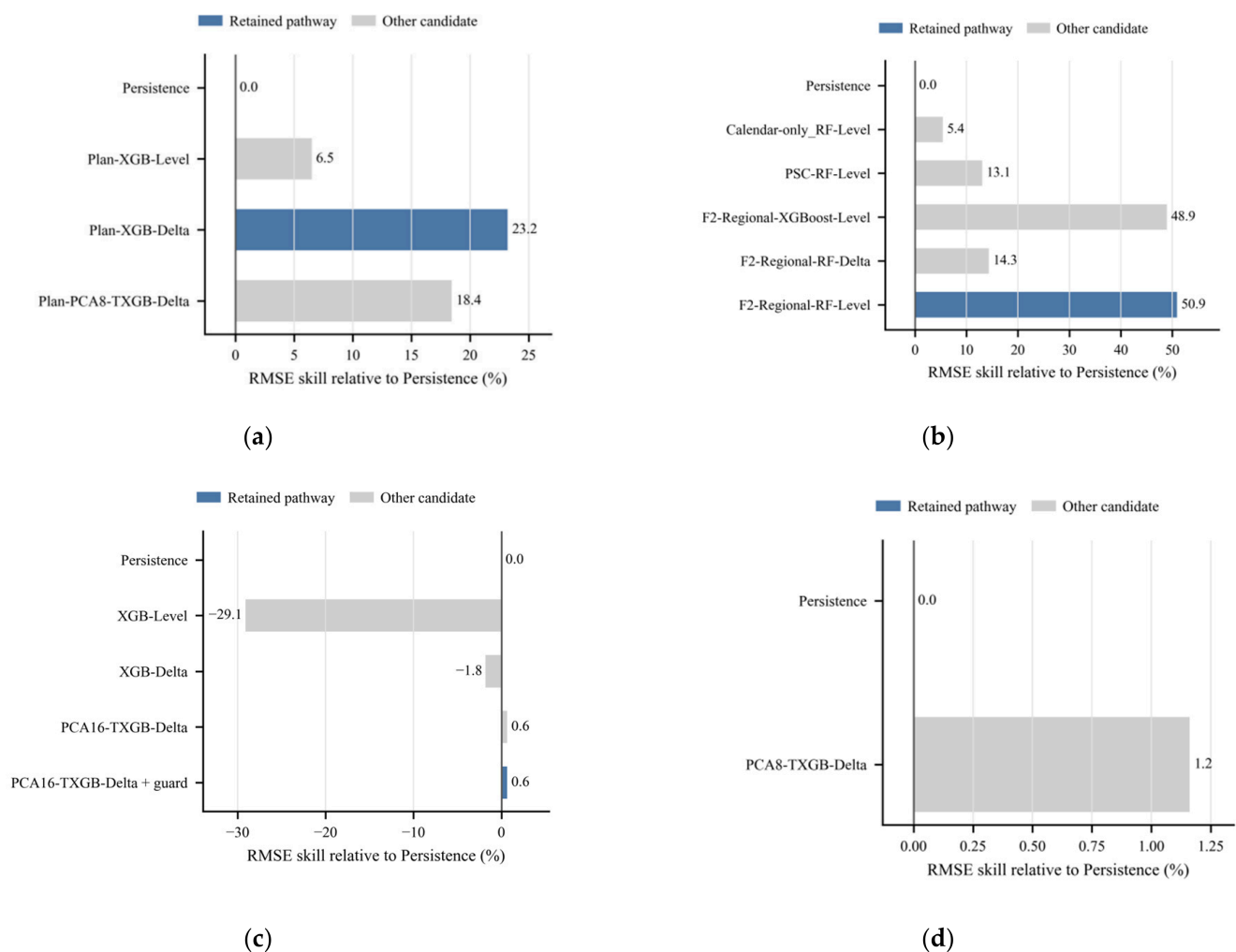


Figure 5. RMSE skill of the pre-specified candidate pathways relative to Persistence over the subsequent 97-day period: (a) local thermal generation; (b) renewable generation; (c) hydropower generation; and (d) imported electricity.

To broaden the benchmark comparison during the post97 stage, TimeXer and PatchTST were included alongside Persistence and the pre-specified model branches and evaluated over the same 97 fully subsequent dates. For thermal, renewable, and hydropower forecasting, the frozen pathways achieved lower RMSE than both contemporary baselines. Relative to TimeXer, the RMSE reductions were 20.5%, 44.3%, and 8.9%, respectively, while the corresponding reductions relative to PatchTST were 27.1%, 45.2%, and 8.8%. Across these six comparisons, the 95% confidence intervals of the RMSE differences, obtained using a 7-day moving-block bootstrap, did not cross zero, indicating that the observed performance advantages remained statistically stable after accounting for temporal dependence. For imported-electricity forecasting, TimeXer achieved a point RMSE 3.4% lower than that of the frozen Persistence pathway, whereas PatchTST produced a higher point RMSE. In both paired comparisons, however, the 95% confidence intervals crossed zero. This indicates that the RMSE difference changed signs across bootstrap resamples and therefore did not provide stable evidence that either contemporary baseline consistently outperformed or underperformed the frozen Persistence pathway over the fully subsequent 97-day period. Accordingly, Persistence, TimeXer, and PatchTST are treated as statistically comparable for imported-electricity forecasting in this evaluation. Table 9 summarizes the direct post97 comparisons.

Table 9. Comparison of the frozen pathways with Persistence and contemporary forecasting baselines over the fully subsequent 97-day period.

Target	Frozen Pathway RMSE (MWh)	Persistence RMSE (MWh)	TimeXer RMSE (MWh)	PatchTST RMSE (MWh)	TimeXer – Frozen Δ RMSE [95% CI]	PatchTST – Frozen Δ RMSE [95% CI]
Thermal	148.645	193.546	187.046	203.793	+38.401 [14.932, 57.740]	+55.148 [22.938, 82.160]
Renewable	2759.543	5621.809	4952.130	5032.484	+2192.587 [1072.903, 3222.544]	+2272.942 [1102.578, 3370.312]
Hydropower	429.113	431.917	471.250	470.520	+42.137 [14.779, 70.551]	+41.406 [7.255, 77.036]
Imported electricity	1728.416	1728.416	1671.586	1800.952	−56.830 [−131.874, 10.050]	+72.536 [−37.269, 160.016]

Note: All methods are evaluated on the same fully subsequent 97 dates. Δ RMSE is defined as baseline RMSE minus frozen-pathway RMSE, so positive values favor the frozen pathway. The paired 95% intervals were obtained using a 7-day moving-block bootstrap with 5000 repetitions. TimeXer and PatchTST were evaluated as post-configuration contemporary reference baselines and did not participate in the original pathway-selection process.

3.4. Training-History Sensitivity and External Pathway Stability

3.4.1. Four-Target Training-History Reduction

The selected four-target configurations were refitted under FULL, N450, N365, and N270 and evaluated on the same 97 dates. Holding the evaluation dates fixed isolates the effect of training-history length. Table 10 and Figure A1 (Appendix A) show that the response is target-dependent and not uniformly monotonic.

Renewable generation was the most sample-sensitive learned pathway: N365 remained within 4.75% of FULL, whereas N270 increased RMSE by 21.61%. Its R^2 decreased from 0.7337 under FULL to 0.7078 at N365 and 0.6061 at N270. Hydropower R^2 declined gradually from 0.7782 to 0.7453, while the thermal N270 setting increased R^2 from 0.8084 to 0.8229. These results show that sample-depth effects are target-dependent and not uniformly monotonic.

Table 10. RMSE changes and corresponding R^2 values under controlled reductions in four-target training-history length.

Target	FULL RMSE/ R^2	N450 Change/ R^2	N365 Change/ R^2	N270 Change/ R^2	Interpretation
Local thermal generation	148.645 MWh $R^2 = 0.8084$	+7.45% $R^2 = 0.7788$	+6.13% $R^2 = 0.7841$	−3.87% $R^2 = 0.8229$	Non-monotonic; moderate sensitivity around one year
Renewable generation	2759.543 MWh $R^2 = 0.7337$	+7.65% $R^2 = 0.6914$	+4.75% $R^2 = 0.7078$	+21.61% $R^2 = 0.6061$	Moderate at N365; marked degradation at N270
Hydropower generation	429.113 MWh $R^2 = 0.7782$	+2.89% $R^2 = 0.7652$	+3.85% $R^2 = 0.7608$	+7.16% $R^2 = 0.7453$	Gradual absolute degradation; weak incremental gain
Imported electricity	1728.416 MWh $R^2 = 0.5212$	0.00% $R^2 = 0.5212$	0.00% $R^2 = 0.5212$	0.00% $R^2 = 0.5212$	Unchanged because Persistence requires no fitted training sample

Note: FULL uses all admissible training history, whereas N450, N365, and N270 retain the most recent 450, 365, and 270 training days. All settings use the same 97 evaluation dates. A negative RMSE change denotes lower RMSE than FULL. R^2 is calculated on the same fixed evaluation dates for each training-history setting. Because Persistence is model free, the imported-electricity predictions and R^2 values are unchanged across these training-depth labels.

3.4.2. External Rolling-Origin Pathway Stability with Retrospective Target-Day Weather

The main external rolling-origin analysis focuses on renewable, solar, and wind generation under retrospectively supplied target-day weather. Table 11 lists the pathway selected in each annual fold and the configuration selected from 2022 evidence and held fixed for 2023. The complete forecast-origin-safe pathways for all six external targets are provided in Supplementary Table S11.

Table 11. Configuration frequencies and 2023 fixed pathways in the retrospective target-day weather evaluation.

Target	Configuration Frequency Across Five Folds	Path Fixed for 2023
Renewable	XGBoost-Level: 3; TXGB-Level: 1; TXGB-Delta: 1	C4-W2-RTE-CapF-TXGB-Delta
Solar	XGBoost-Level: 2; XGBoost-Delta: 1; TXGB-Level: 1; TXGB-Delta: 1	C1-W2-RTE-CapF-XGBoost-Level
Wind	XGBoost-Level: 2; XGBoost-Delta: 1; TXGB-Delta: 2	C4-W2-RTE-CapF-TXGB-Delta

Note: The five selections comprise the four rolling folds for 2019–2022 and the pathway selected from 2022 with inner evidence for the fixed 2023 evaluation. Complete fold-specific pathway labels are reported in Supplementary Table S12.

Under the forecast-origin-safe external registry, FULL R^2 values for fossil thermal generation, renewable generation, hydropower generation, net physical exchange, solar generation, and wind generation were 0.8743, 0.4660, 0.9007, 0.8371, 0.8520, and 0.5027, respectively. These results define the operationally admissible information boundary, whereas the W2-RTE results reported below represent a retrospective complete-weather benchmark.

The selected target representation and predictor family varied across folds, although all candidates shared target-day weather and previous-year capacity normalization. This variability supports fold-specific configuration rather than a single fixed model for all years and targets.

3.4.3. External Performance and 365-Day Training-History Sensitivity

Table 12 reports the external rolling-origin performance of renewable, solar, and wind generation under retrospectively supplied target-day weather. With the full training history,

R^2 reached 0.9383, 0.9706, and 0.9509, respectively. Restricting training to the most recent 365 days increased RMSE by 34.89%, 28.39%, and 32.63%. Forecast-origin-safe FULL and recoverable N365 results are reported separately in Supplementary Table S13.

Table 12. External retrospective target-day weather performance and N365 sensitivity over the 2019–2023 outer folds.

Target	FULL RMSE (MWh)	FULL R^2	N365 RMSE (MWh)	N365 R^2	RMSE Change
Renewable	18,173.759	0.9383	24,514.593	0.8878	+34.89%
Solar	4066.762	0.9706	5221.180	0.9515	+28.39%
Wind	17,216.579	0.9509	22,833.842	0.9136	+32.63%

Note: FULL and N365 use the same 1826 aligned outer evaluation dates. The target-day weather inputs are retrospective and define a retrospective complete-weather benchmark and not an operational weather feed.

The three weather-driven targets remained accurate under the full-history setting but degraded when training was limited to the most recent 365 days. The target-specific magnitude of this decline indicates that training-history requirements depend on the selected information and representation pathway.

3.5. Fixed-Path 2023 Evaluation Under Retrospective Target-Day Weather

Using 2022 as the final offline configuration period, one retrospective target-day weather pathway was selected for each of renewable, solar, and wind generation, refitted with admissible data available before 2023, and held fixed throughout 2023. The 2023 target observations were fully excluded from pathway selection and model fitting; however, target-day ERA5 covariates were supplied retrospectively. This experiment therefore represents a held-out retrospective-weather benchmark rather than an operational weather-forecast evaluation. Table 13 reports the fixed-path 2023 results. Forecast-origin-safe results for all six external targets are retained in Supplementary Table S14.

Table 13. Fixed-path 2023 performance under retrospective target-day weather.

Target	2022-Selected Frozen Path	2023 RMSE (MWh)	2023 R^2
Renewable	C4-W2-RTE-CapF-TXGB-Delta	22,794.939	0.9246
Solar	C1-W2-RTE-CapF-XGBoost-Level	5598.011	0.9621
Wind	C4-W2-RTE-CapF-TXGB-Delta	22,055.286	0.9456

Note: All pathways were selected using 2022 inner evidence, refitted with admissible data available before 2023, and held fixed throughout 2023. Target-day weather was supplied retrospectively.

The fixed 2023 pathways achieved R^2 values of 0.9246, 0.9621, and 0.9456 for renewable, solar, and wind generation, respectively.

The fixed-path 2023 results were consistent with the rolling-origin evaluation and showed that the selected pathways remained accurate over a complete held-out target year under the same retrospective target-day weather setting.

3.6. Controlled Information Unavailability

To examine incomplete operational information, the capacity-normalized candidate family was removed before fold-specific selection, after which configuration and refitting were repeated. For renewable, solar, and wind generation, this removal also excluded the retrospective target-day weather pathways. Table 14 reports the resulting changes. The separate historical-weather-interface removal is provided in Supplementary Table S5. This experiment therefore characterizes candidate-space availability and the resulting

pathway reconfiguration, while the component-specific effects of capacity scaling, W2-RTE availability, target representation, and predictor family are evaluated separately in the decoupled factorial ablation.

Table 14. Performance after removal of the capacity-normalized candidate family.

Target	Changed Folds	Pooled RMSE (MWh)	R ²	RMSE Change
Renewable	5/5	53,407.994	0.4674	+193.874%
Solar	5/5	9091.208	0.8528	+123.549%
Wind	5/5	54,804.567	0.5024	+218.324%

Note: All five folds selected a different pathway after the candidate-family removal. Complete reconfigured paths are reported in Supplementary Table S15. Because the removal also excluded the retrospective target-day weather pathways, the changes represent the combined effect of candidate removal, fold-specific reconfiguration, and refitting. The reported values characterize the combined candidate-space availability and reconfiguration response. Component-specific effects are reported separately in the decoupled factorial ablation.

Removing the capacity-normalized candidate family changed the selected pathway in all five folds for renewable, solar, and wind generation and increased pooled RMSE by 193.874%, 123.549%, and 218.324%, respectively. These values characterize the combined candidate-space reconfiguration and refitting response, while the component-specific effects are reported separately in Section 3.7.

3.7. Decoupled Factorial Ablation of Pathway Components

The decoupled factorial analysis separated the individual effects of capacity normalization, retrospective target-day weather availability, target representation, and predictor family under matched conditions. For each factor, only the factor of interest was changed, while the other three factors were kept unchanged. Table 15 summarizes the pooled matched RMSE differences and the corresponding 95% moving-block-bootstrap intervals for the external renewable, solar, and wind targets.

Table 15. Matched RMSE effects of the four-factor decoupled ablation for the external RTE weather-driven targets.

Factor	Matched Contrast (A – B)	Renewable Δ RMSE (MWh) [95% CI]	Solar Δ RMSE (MWh) [95% CI]	Wind Δ RMSE (MWh) [95% CI]
Capacity normalization	Raw – CapF	+3680.167 [2870.879, 4488.769]	+1676.214 [1352.246, 2025.833]	+1967.901 [1392.603, 2577.095]
Target-day weather	no-W2 – W2-RTE	+35,551.986 [33,160.564, 37,921.402]	+3477.987 [3195.771, 3765.580]	+37,576.828 [35,042.375, 39,998.338]
Target representation	Delta – Level	–1381.543 [–2322.157, –480.318]	–1241.283 [–1604.485, –903.358]	–1129.588 [–2204.716, –155.258]
Predictor family	XGB – TXGB	–4231.384 [–5168.995, –3350.233]	+180.560 [21.665, 338.542]	–4924.573 [–5866.091, –3998.227]

Note: Δ RMSE = RMSE(A) – RMSE(B) for the matched contrast shown in the second column. Positive values indicate lower RMSE for condition B, whereas negative values indicate lower RMSE for condition A. Each comparison changes only the indicated factor while holding the other three factors fixed. The 95% intervals were obtained using a 7-day moving-block bootstrap with 5000 repetitions. W2-RTE denotes retrospective target-day ERA5 reanalysis information.

The matched factorial analysis reveals clear differences in the contributions of individual pathway components. Retrospective target-day W2-RTE availability produced the largest RMSE reductions, particularly for renewable and wind generation, while capacity normalization provided consistent improvements across all three external targets. The

preferred target representation in these external comparisons was Delta, and the predictor-family effect varied by target, with XGBoost performing better for renewable and wind generation and TXGB showing a smaller advantage for solar generation. Taken together, these results show that information availability, target representation, and predictor family contribute at different magnitudes across forecasting targets, supporting their treatment as distinct target-dependent dimensions within pathway configuration.

3.8. Configuration Effects and Target-Specific Model Interpretation

Table 16 summarizes the performance differences associated with the main configuration choices examined in this study. It includes comparisons of the renewable information interface, target representation, and predictor family; the hydropower temporal representation and output rule; and external candidate-family availability. The comparisons were conducted under their respective chronological protocols, and each result is interpreted within the corresponding evaluation setting.

Table 16. Controlled pathway and candidate-space analyses across target-specific evaluation records.

Analysis Dimension	Controlled Comparison and Evaluation Record	Main Result	Interpretation
Information interface	F2-Regional vs. PSC for renewable generation; pooled T3–T5 dates	F2-Regional was associated with a 1068.633 MWh lower RMSE.	Regional weather information was a major source of the renewable-performance difference.
Target representation	Level vs. Delta for renewable generation; pooled T3–T5 dates	Level was associated with a 1292.274 MWh lower RMSE.	The preferred target representation depended on the operating characteristics of the target.
Predictor family	RF vs. XGBoost for renewable generation; pooled T3–T5 dates	RF was associated with a 204.894 MWh lower RMSE.	The predictor-family difference was smaller than the information-interface and representation differences.
Temporal representation	PCA16-TXGB-Delta vs. no-latent XGB-Delta for hydropower; post97	PCA16-TXGB-Delta achieved a 2.44% lower RMSE.	The compressed temporal representation provided a limited target-specific gain.
Output rule	Pre-guard vs. guarded hydropower forecasts; post97	The guard remained inactive, and the two forecasts were identical.	The guard served as an output safety rule during this period.
Candidate-family availability	Complete candidate space vs. removal of the CapF family and associated retrospective W2-RTE pathways; external rolling-origin folds	All folds were reconfigured, and pooled RMSE increased by 123.55–218.32%.	Candidate-family availability affected both pathway selection and subsequent performance.

Note: Each row reports an independent controlled analysis using the target and evaluation record specified in the second column. Because the rows differ in target, period, and result scale, the reported values are interpreted only within their respective comparisons. In the final row, the RMSE change represents the overall performance response after removing the capacity-normalized pathway set and its associated retrospective W2-RTE pathways, followed by fold-specific pathway reselection and model refitting.

Across the renewable comparisons, the differences associated with the information interface and target representation were larger than the predictor-family difference. For hydropower, the temporal representation produced a smaller improvement, while the guard acted as a safety rule during post97. The external candidate-family analysis fur-

ther shows that changes in pathway eligibility can affect both fold-specific selection and the performance obtained after refitting. Figure 6 complements these comparisons with interpretation evidence matched to each retained forecasting mechanism.

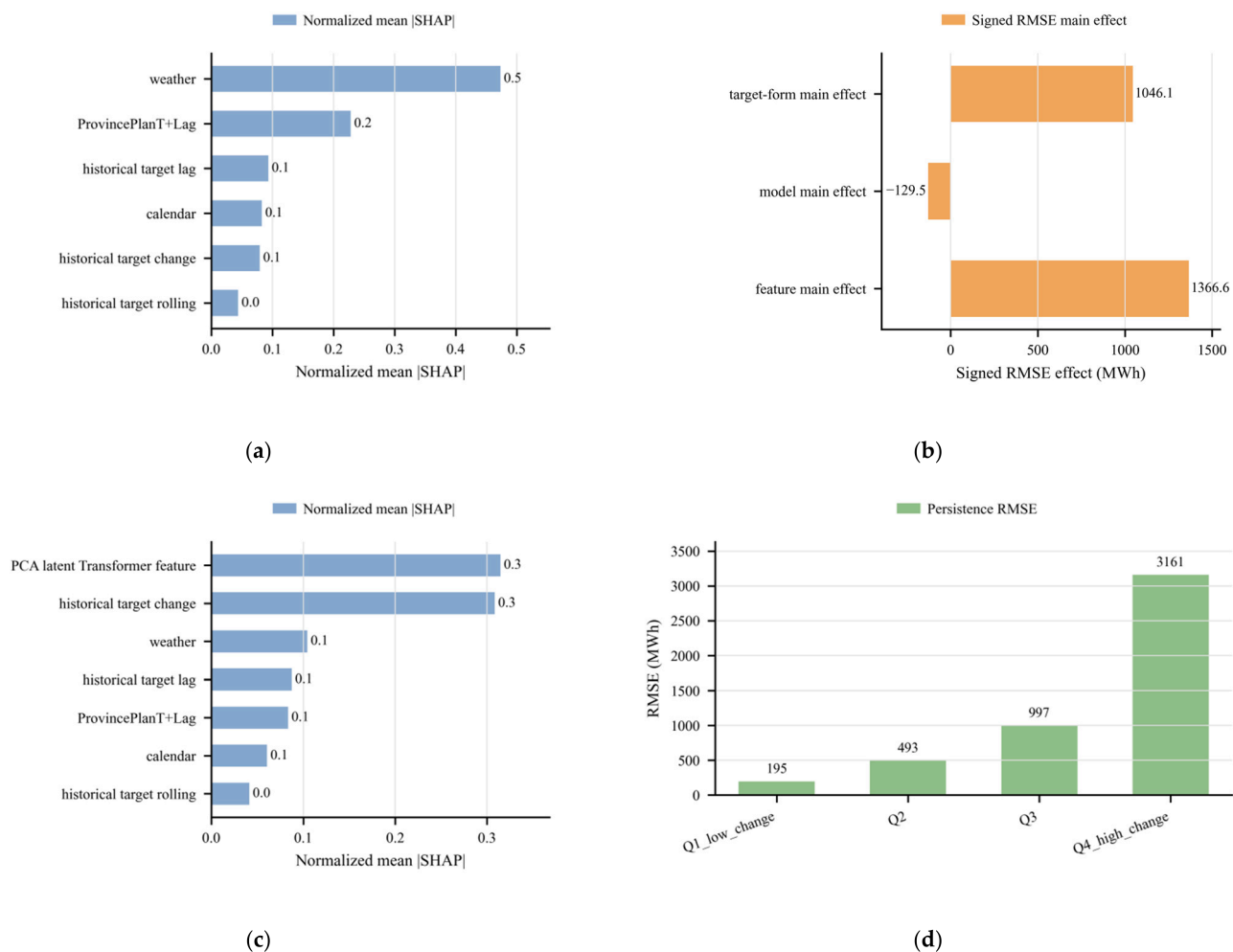


Figure 6. Target-specific interpretation of the retained forecasting pathways: (a) normalized mean absolute SHAP contributions for local thermal generation; (b) signed post97 RMSE main effects for renewable generation; (c) normalized mean absolute SHAP contributions for hydropower generation; and (d) Persistence RMSE across quartiles of absolute inter-day target change for imported electricity. Positive values in panel (b) favor F2-Regional over PSC and Level over Delta, whereas negative values favor XGBoost over RF.

Figure 6 provides target-appropriate interpretation of the retained forecasting mechanisms. For local thermal generation, the fitted model relied most strongly on weather variables and ProvincePlanT+Lag according to mean absolute SHAP values. For renewable generation, panel (b) reports signed post97 main effects: positive values indicate lower RMSE for F2-Regional than PSC and for Level than Delta, whereas the negative model effect indicates lower RMSE for XGBoost than RF during post97. Across the pooled T3–T5 common-date record, F2-Regional, Level, and RF were associated with RMSE advantages of 1068.633, 1292.274, and 204.894 MWh, respectively.

For hydropower, the fitted XGBoost core relied most strongly on the PCA-compressed Transformer features and historical target change according to normalized TreeSHAP values. The post hoc 97-day comparison was associated with a 2.44% lower RMSE for the latent pathway than for the no-latent XGB-Delta branch. The output guard was inactive, so the pre-guard and final forecasts were identical. For imported electricity, Persistence errors

increased across target-change quantiles, indicating that level continuity was informative while large daily changes remained difficult to anticipate.

3.9. Chronological Rolling and Adaptive Uncertainty Performance

The fixed, rolling, and adaptive interval-calibration schemes were evaluated over the same fully subsequent 97-day post97 period for renewable and hydropower generation. The frozen point forecasts were identical across FIXED, RC60, and ACI60, and only the interval-calibration procedure varied. Renewable calibration followed the date-balanced convention defined in Section 2.9.3, aggregating 533 chronological out-of-sample residual records to 357 unique calibration dates. Figure 7 shows the nominal 90% prediction intervals obtained using the FIXED calibration scheme for renewable and hydropower generation during the subsequent 97-day period. For each target, the interval half-width is calibrated using pre-post97 chronological out-of-sample absolute residuals and remains fixed throughout this period.

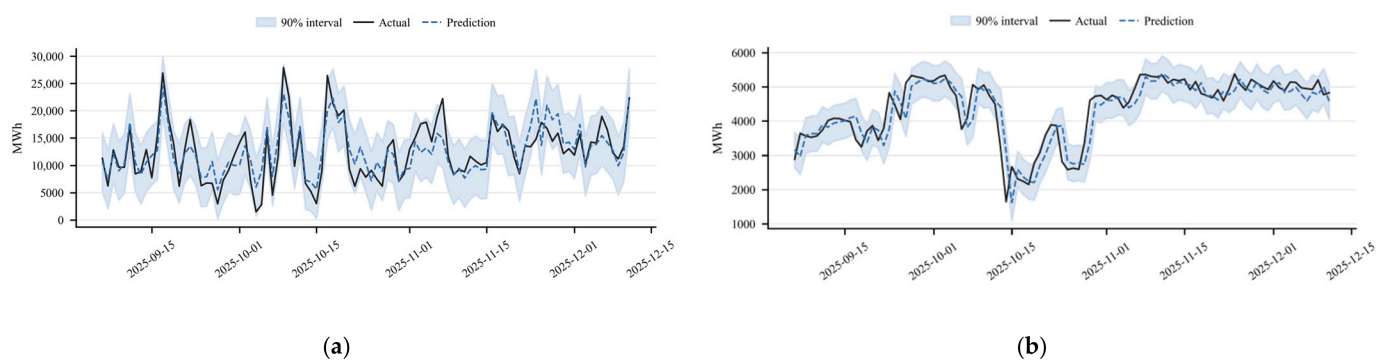


Figure 7. Nominal 90% prediction intervals obtained using the FIXED calibration scheme during the fully subsequent 97-day post97 period: (a) renewable generation and (b) hydropower generation.

As shown in Table 17, the FIXED renewable interval yielded $PICP = 0.9588$ and $MPIW = 10,715.370$ MWh. RC60 brought the coverage closer to the nominal 0.90 level while reducing $MPIW$ to 9315.210 MWh, and ACI60 produced the same $PICP$ of 0.9175 with $MPIW = 9310.011$ MWh and $MIS = 11,406.530$. For hydropower, FIXED yielded $PICP = 0.8454$. RC60 increased coverage to 0.8660, and ACI60 further increased it to 0.8969, corresponding to a coverage error of 0.0031. The accompanying increase in hydropower interval width reflects the coverage–sharpness trade-off under sequential adaptation.

Table 17. Performance of fixed, rolling, and adaptive 90% prediction intervals over the fully subsequent post97 period.

Target	Method	PICP	Coverage Error	MPIW (MWh)	MIS
Renewable generation	FIXED	0.9588	0.0588	10,715.370	11,908.540
	RC60	0.9175	0.0175	9315.210	11,408.794
	ACI60	0.9175	0.0175	9310.011	11,406.530
Hydropower generation	FIXED	0.8454	0.0546	1036.685	2273.374
	RC60	0.8660	0.0340	1623.692	2398.068
	ACI60	0.8969	0.0031	1840.424	2358.170

Note: The nominal coverage level is 0.90, and coverage error is defined as $|PICP - 0.90|$. FIXED uses the pre-post97 calibration residuals with a constant interval width throughout post97. RC60 recalibrates the interval from the most recent 60 admissible residual dates, while ACI60 further updates the effective miscoverage level after each newly observed outcome. The frozen point forecasts are identical across all three schemes.

Following the primary FIXED–RC60–ACI60 comparison, the robustness analysis was further extended to examine directional asymmetry in forecast errors and forecast-magnitude dependence. The asymmetric and magnitude-normalized variants characterize residual direction and forecast scale, respectively. For renewable generation, Asym-RC60 achieved $\text{PICP} = 0.8969$ with $\text{MPIW} = 8946.812$ MWh, while Mag-ACI60 achieved $\text{PICP} = 0.9072$. These results show that incorporating residual direction or forecast scale can alter the balance between empirical coverage and interval width, with the resulting effects depending on the calibration scheme and forecasting target. The complete asymmetric and magnitude-normalized results for renewable and hydropower generation are reported in Supplementary Table S6, and dependence-aware MIS comparisons based on a 7-day moving-block bootstrap are provided in Supplementary Table S7.

The weather-conditioned diagnostic further showed that renewable interval coverage varied across 100 m wind-speed and shortwave-radiation regimes. Using fixed stratification thresholds determined from the pre-post97 calibration data, FIXED, RC60, and ACI60 exhibited different degrees of Low/High coverage variation across both weather dimensions, indicating that interval behavior remained conditioned on meteorological states beyond the aggregate coverage level. Detailed regime definitions and stratified results are reported in Supplementary Table S8. Because the High-radiation regime contained relatively few observations, the corresponding results are used primarily to characterize interval coverage under that condition and to inform future conditional calibration over longer evaluation periods.

Overall, the uncertainty analyses show target-dependent interval behavior across sequential updating and the complementary robustness dimensions. For renewable generation, rolling and adaptive updating reduced interval width while retaining coverage close to the nominal level. For hydropower, adaptive calibration brought empirical coverage close to 0.90 with a wider interval, reflecting a different coverage-sharpness balance between the two targets.

4. Discussion

The results show that the retained pathways were shaped primarily by target-specific information and representation choices. The thermal model combined plant-plan context with Delta prediction, whereas the renewable model used the regional target-day weather-forecast interface with a Level target. In the controlled renewable comparison, differences associated with information-interface choices and the choice between Level and Delta target representations were larger than the performance difference between RF and XGBoost. The fitted hydropower model relied on PCA-compressed temporal features and Delta reconstruction, while the guard only constrained its output. For imported electricity, Persistence was retained given the available historical and calendar information.

These choices were consistent with the mechanisms of each target. The Level representation preserved the direct relationship between weather conditions and renewable output, whereas Delta emphasized short-term changes. The modest gain from latent features for hydropower suggests that temporal encoding could not compensate for the absence of streamflow and reservoir-state data. Retaining Persistence for imported electricity likewise reflects the absence of transaction schedules, tie-line plans, and external dispatch states.

The chronological evidence layers serve distinct purposes. The 60-day period determines the internal pathways, fixed95 provides a common-date developmental comparison, and T3–T5 provide retrospective diagnostics. Although T4 overlaps the configuration and fixed95 periods and T5 falls within fixed95, neither window is used to reselect or modify the frozen pathways. The following 97-day period provides a fully subsequent evaluation. For renewable generation, XGBoost achieved an RMSE 58.441 MWh lower

than that of the retained RF pathway during formal60. However, the paired-date bootstrap interval in Table 4 included zero, and the estimated probability of improvement was 0.7336, below the pre-specified replacement threshold of 0.95. The criterion therefore avoids replacing the retained pathway on the basis of a modest and statistically uncertain short-window difference.

The repeated-window analysis provides additional evidence on the stability of the formal60 decisions. The candidate with the lowest point RMSE changed across some 60-day windows, showing that rankings were moderately sensitive to window placement. Nevertheless, the pre-specified rule retained the original pathways in all eight thermal windows, six of eight renewable windows, six of eight hydropower windows, and seven of eight imported-electricity windows. After the 7-day moving-block bootstrap and Holm correction, evidence favoring an alternative occurred in none of the thermal or renewable windows and in only one hydropower and one imported-electricity window. No such evidence appeared for any target in the non-overlapping subset. Hydropower was the most sensitive target, but its median regret relative to the lowest-RMSE candidate was only 5.941 MWh, and the single replacement signal did not persist in the non-overlapping analysis. Thus, point rankings varied, but multiplicity-controlled evidence for changing the formal60 decisions remained limited.

The effect of training-history length also depended on the information boundary. Under retrospective target-day weather, renewable, solar, and wind achieved R^2 values of 0.9383, 0.9706, and 0.9509, respectively, but all three deteriorated when the training history was limited to 365 days. The forecast-origin-safe historical-weather results in Supplementary Table S13 provide a boundary comparison rather than operational support for the retrospective-weather findings. Supplementary Table S15 shows that removing the capacity-normalized family altered both candidate eligibility and the refitted pathways. The resulting performance change therefore reflects candidate-space reconfiguration rather than the isolated effect of a single component.

The decoupled factorial analysis clarifies the candidate-family removal experiment in Table 14. The large RMSE increases reflect the combined effects of excluding capacity-normalized pathways, losing W2-RTE pathway availability, and reconfiguring the candidate space within each fold. Under matched conditions, retrospective target-day W2-RTE availability produced the largest performance gains, while CapF consistently improved performance across the three external targets. Delta was favored in the matched representation comparisons, whereas the preferred predictor family varied by target. These results distinguish the target-dependent contributions of information availability, target representation, and predictor family and support treating them as separate dimensions of pathway configuration.

The comparison with TimeXer and PatchTST indicates that forecasting performance under limited samples and constrained information availability depends jointly on the predictor, information interface, and target representation. The frozen target-specific pathways achieved statistically supported RMSE reductions relative to both contemporary baselines for thermal, renewable, and hydropower generation. For imported electricity, no statistically supported performance differences were detected among Persistence, TimeXer, and PatchTST over the fully subsequent 97-day period. This result is consistent with the limited explanatory information available for imported electricity and reinforces the importance of matching the predictor to the target-specific information interface and representation.

The extended uncertainty analysis shows that interval calibration is target-dependent. For renewable generation, RC60 and ACI60 reduced interval width while maintaining empirical coverage close to the nominal level. For hydropower, ACI60 increased PICP from 0.8454 to 0.8969, although the wider interval revealed a clear coverage-sharpness

trade-off. Additional diagnostics identified directional error asymmetry, variation with forecast magnitude, and differences in renewable coverage across wind-speed and radiation regimes. Aggregate post97 coverage therefore did not fully represent conditional interval behavior. This regime dependence motivates calibration over longer horizons once sufficiently large regime-specific samples become available. Overall, chronological recalibration can accommodate changing forecast-error conditions, but the preferred scheme remains target-dependent.

The history-reduction experiment demonstrates that the effect of training-history length is target-specific and non-monotonic. For thermal generation, N270 outperformed FULL, whereas N365 and N450 performed worse. This pattern is consistent with differences in the operating stages and sample compositions represented by each window, including possible shifts in operating baselines or dispatch policies. FULL provides broader coverage of recurring operating conditions and long-term variation, whereas N270 is more closely aligned with the evaluation period and may better represent recent operations. N365 and N450 provide intermediate combinations of historical coverage and recency, resulting in different predictive outcomes. For renewable generation, performance deteriorated consistently as the training history shortened, particularly under N270, indicating that longer records are important for capturing seasonal and weather-related variation. Hydropower performance also declined progressively with shorter histories, indicating greater dependence on long-term operational information. Training-history selection should therefore balance historical coverage with relevance to recent operating conditions. In practice, validation-guided rolling-window updating offers a practical means of achieving this balance.

For imported electricity, Persistence achieved an R^2 of 0.5212. This result indicates that recent exchange history is informative but insufficient to represent operational arrangements that are absent from the available inputs. Interregional dispatch arrangements, tie-line schedules, transaction plans, and transmission constraints could help explain future exchange variation. Incorporating these variables may enable forecasting pathways to outperform the naïve baseline.

The empirical residual intervals also showed target-specific calibration behavior. Date-balanced calibration covered 93 of 97 post97 days for renewable generation, and duplicate-date handling had little effect. Hydropower intervals covered 82 of 97 days, below the nominal 90% level. This shortfall indicates that the fixed absolute-residual intervals did not fully accommodate the subsequent hydropower error distribution. Practical deployment should therefore retain forecast issue times, publication-dated capacity records, explicit plan-variable definitions, and continued target-specific interval monitoring.

Operationally, thermal and renewable forecasts can inform day-ahead volume assessment, bidding, and supply balancing. Operational use of the hydropower pathway requires continued monitoring of incremental performance gains and interval coverage. Persistence should remain the benchmark for imported electricity until transaction schedules, tie-line plans, and external dispatch states become available.

The generalizability of these findings is constrained by the scope and information conditions of the analyses. The internal study covers a single regional power system; F2-Regional was reconstructed retrospectively from archived forecast runs; and W2-RTE uses target-day reanalysis. Pathway configurations may therefore differ across systems and forecast-origin information settings. Hydropower achieved coverage close to the nominal level only with a wider adaptive interval, highlighting the need to update interval calibration as forecast-error characteristics evolve. Future applications should reassess pathway eligibility and interval calibration using the information available at each forecast origin.

5. Conclusions

This study proposes a Delta-enhanced, validation-guided framework for target-specific daily electricity-supply forecasting under limited samples and incomplete operational information. The framework jointly configures forecast-origin-safe information interfaces, target representations, predictors, temporal components, and output rules. Chronological validation retained a plan-informed Delta pathway for thermal generation, an archived-weather-informed Level pathway for renewable generation, a PCA-enhanced TXGB-Delta pathway with a persistence-based output guard for hydropower, and Persistence for imported electricity under the available historical and calendar information. Each pathway was refitted using admissible pre-evaluation data and then frozen for subsequent forecasting.

The fully subsequent post97 evaluation confirmed the performance of the four internal pathways, with R^2 values ranging from 0.5212 to 0.8084. Relative to Persistence, the thermal and renewable pathways reduced RMSE by 23.2% and 50.9%, respectively. In the external retrospective evaluation, the three weather-driven targets achieved R^2 values ranging from 0.9383 to 0.9706, but performance remained sensitive to training-history length and candidate-family availability. Residual intervals covered 93 of 97 subsequent days for renewable generation and 82 of 97 days for hydropower, underscoring the need for target-specific calibration. Under the stated forecast-origin boundaries, the internal results support target-specific forecasting for day-ahead electricity-volume assessment, market bidding, regional supply balancing, and interregional procurement, while the external analyses provide complementary sensitivity evidence. Future work will incorporate publication-timestamped operating information, archived numerical weather forecasts for external systems, and longer-horizon interval calibration that accounts for seasonal and regime-dependent variations.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/en19184248/s1>, Table S1: Exact construction of the Province-PlanT+Lag information vector; Table S2: Exact construction of the 13-dimensional PSC vector; Table S3: Implemented Open-Meteo fixed-24-h-lead weather features used in the internal regional renewable-generation interface; Table S4: ERA5-derived external weather-interface variables and temporal aggregation rules; Table S5: Fold-specific reconfiguration after removal of the historical-weather interface; Table S6: Robustness checks using asymmetric and forecast-magnitude-normalized 90% interval calibration; Table S7: Selected 7-day moving-block-bootstrap comparisons of mean interval score; Table S8: Weather-regime-conditioned performance of the renewable 90% prediction intervals during post97; Table S9: Documented reproducibility settings and roles of the internal and external candidate families; Table S10: Main internal candidate families and target-specific eligibility in the 60-day configuration period; Table S11: Forecast-origin-safe fold-specific pathway selections for the six external targets; Table S12: Complete fold-specific pathway selections in the retrospective target-day weather evaluation; Table S13: Forecast-origin-safe FULL performance and available N365 sensitivity over the 2019–2023 outer folds; Table S14: Forecast-origin-safe fixed-path performance during the held-out 2023 evaluation; Table S15: Complete fold-specific reconfiguration after removal of the capacity-normalized candidate family; Algorithm S1: Chronological generation of PSC states.

Author Contributions: Conceptualization, L.T. and Y.C.; methodology, Y.C.; software, Y.C.; validation, Y.C.; formal analysis, Y.C.; investigation, Y.C., L.Z. and Z.Y.; resources, L.T.; data curation, L.T. and Y.C.; writing—original draft preparation, Y.C.; writing—review and editing, L.T. and Y.C.; visualization, Y.C. and Y.L.; supervision, L.T. and P.H.; project administration, L.T., P.H. and T.L.; funding acquisition, L.T. and P.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Henan Provincial Natural Science Foundation, grant number 252300420471; the Henan Provincial Science and Technology Research Project, grant number 242102240123; and the 2026 Henan Provincial University–Enterprise Collaborative Innovation Project entitled “Research and Application of Transformer Acoustic-Fingerprint Imaging Fault Diagnosis Technology Based on an Acoustic Sensor Array”, grant number 26AXQXT065.

Data Availability Statement: The primary regional dataset contains non-public plant-level and regional operating records and cannot be deposited in an open repository because of data-owner restrictions. De-identified derived tables, feature definitions, model configurations, and implementation code are available from the corresponding author upon reasonable request and subject to approval by the data owner. The external experiments use the publicly available RTE éCO2mix national consolidated and definitive dataset [46], the RTE/ODRÉ annual installed-capacity dataset [47], and ERA5 global reanalysis data [48].

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT (OpenAI, GPT-5.5 and GPT-5.6) to assist with the translation of selected content from Chinese into English, as well as with language polishing and formatting. All AI-assisted text was reviewed and edited by the authors, who take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations and study-specific labels are used in this article:

AdamW	Adam optimizer with decoupled weight decay.
ACI60	60-day adaptive conformal interval calibration.
CapF	Previous-year-capacity-normalized 24 h daily utilization.
C1–C4	Retrospective target-day weather, capacity-normalized external candidate identifiers.
Delta	Inter-day change target representation.
D – 1	The day immediately preceding the forecast target day.
ERA5	Fifth-generation atmospheric reanalysis produced by the European Centre for Medium-Range Weather Forecasts.
F1–F5	Five annual rolling-origin outer folds used in the external RTE evaluation.
F2-Regional	Internal regional target-day weather-forecast interface retrospectively assembled from archived previous-run forecasts.
F2-RTE-Hist	Forecast-origin-safe historical-weather interface for the external RTE data.
fixed95	Aligned 95-day common-window developmental comparison period.
FULL/N450/N365/N270	Training-history settings using the full admissible history or the most recent 450, 365, or 270 training days.
GELU	Gaussian error linear unit.
Level	Direct absolute-output target representation.
LGB	LightGBM.
MAE	Mean absolute error.
MIS	Mean interval score.
MPIW	Mean prediction interval width.
MSE	Mean squared error.
nRMSE	Normalized root mean squared error.
ODRÉ	Open Data Réseaux Énergies.
PCA	Principal component analysis.
PICP	Prediction interval coverage probability.
post97	Fully subsequent 97-day frozen-path confirmation period.
PSC	Predicted-state coupling.
R ²	Coefficient of determination.
RC60	60-day rolling conformal interval calibration.

RF	Random Forest.
RMSE	Root mean squared error.
RTE	Réseau de Transport d'Électricité, the French transmission system operator.
SHAP	SHapley Additive exPlanations.
T3–T5	Three retrospective diagnostic windows used in the internal evidence hierarchy.
TH/RE/HY/IM	Plant-level thermal generation, regional renewable generation, regional hydropower generation, and regional imported-electricity targets, respectively.
TXGB	Transformer/PCA temporal representation followed by XGBoost regression.
W2-RTE	Retrospective target-day ERA5 weather interface for the external RTE data.
XGB	XGBoost.

Appendix A

Appendix Figure A1 presents the training-history sensitivity visualization. Detailed reproducibility information and diagnostic results are provided in Supplementary Tables S1–S15 and Supplementary Algorithm S1.

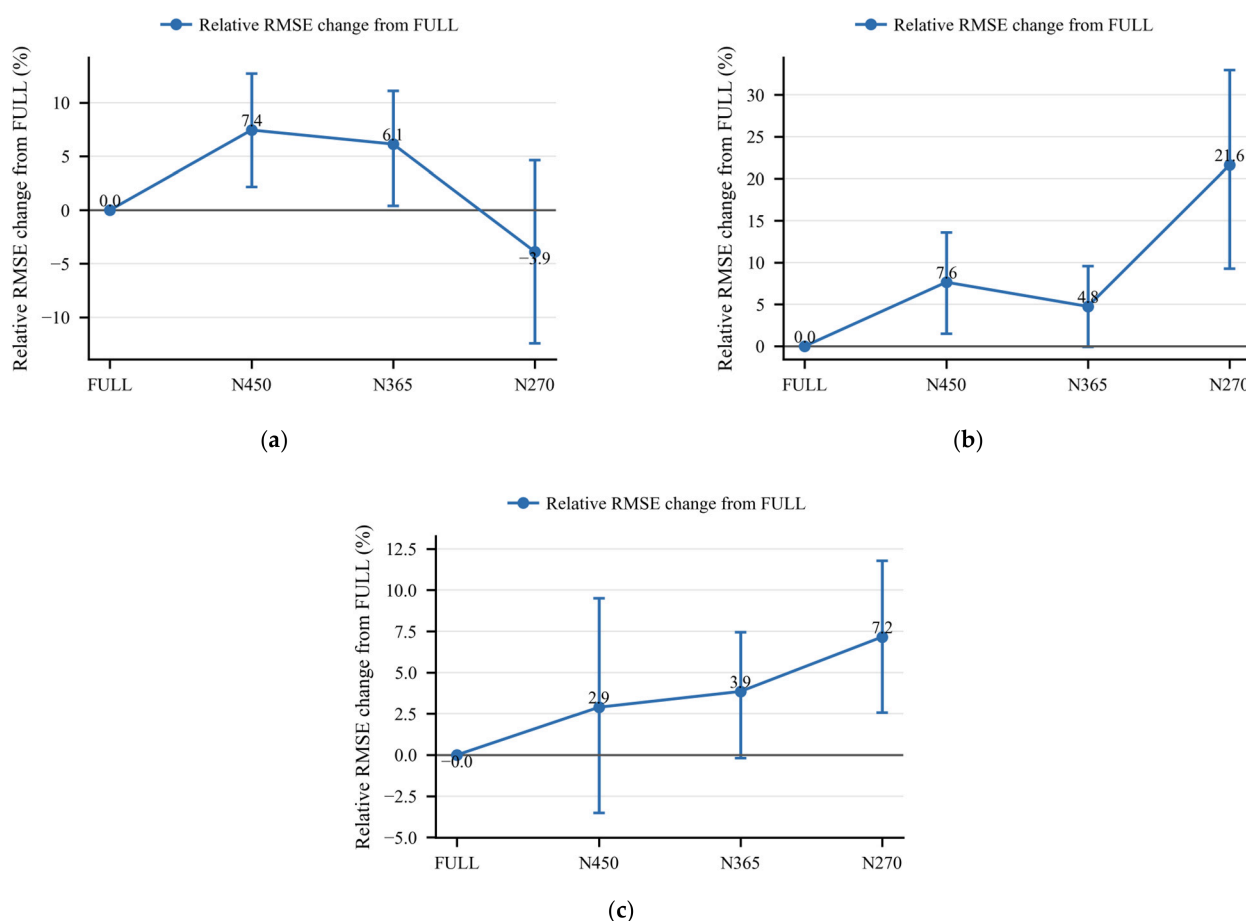


Figure A1. Relative RMSE changes under controlled reductions in four-target training-history length: (a) local thermal generation; (b) renewable generation; and (c) hydropower generation.

References

1. Ma, Z.; Wang, S.; Li, S.; Shi, Y. Long-term coordination for hydro-thermal-wind-solar hybrid energy system of provincial power grid. *Energy Procedia* **2019**, *158*, 6231–6235. [\[CrossRef\]](#)
2. Wang, Y.; Yang, Z.; Li, Y.; Dong, H.; Xing, J. Integrated planning of source-grid-load-storage for regional power grids considering large-scale renewable energy integration. *Energy* **2026**, *345*, 140146. [\[CrossRef\]](#)
3. Basu, M. Fuel constrained commitment scheduling for combined heat and power dispatch incorporating electric vehicle parking lot. *Energy* **2023**, *276*, 127293. [\[CrossRef\]](#)
4. Cai, Q.; Qing, J.; Xue, M.M.; Xu, Q.; Zhong, C.; Liang, Q.M. Techno-economic benefits and energy storage gains of wind–solar complementary power generation: A provincial analysis in China. *Energy Convers. Manag.* **2026**, *349*, 120929. [\[CrossRef\]](#)
5. Carpio, L.G.T.; Guimarães, F.A.C. Regional diversification of hydro, wind, and solar generation potential: A mean-variance model to stabilize power fluctuations in the Brazilian integrated electrical energy transmission and distribution system. *Renew. Energy* **2024**, *235*, 121266. [\[CrossRef\]](#)
6. Wang, Z.; Liu, H.; Wu, S.; Liu, N.; Liu, X.; Hu, Y.; Fu, Y. Explainable time-varying directional representations for photovoltaic power generation forecasting. *J. Clean. Prod.* **2024**, *468*, 143056. [\[CrossRef\]](#)
7. Liu, Y.; Ji, C.; Wang, Y.; Zhang, Y.; Jiang, Z.; Ma, Q.; Hou, X. Effect of the quality of streamflow forecasts on the operation of cascade hydropower stations using stochastic optimization models. *Energy* **2023**, *273*, 127298. [\[CrossRef\]](#)
8. Cai, Q.; Qing, J.; Xu, Q.; Zhong, C.; Xue, M.M.; Liang, Q.M. Incremental techno-economic benefits of interregional generation-grid-load-storage coordination in China's power system decarbonization: A long-term evolutionary and provincial heterogeneity analysis. *Energy Policy* **2026**, *209*, 114995. [\[CrossRef\]](#)
9. Schulte, R.H.; Wahle, R.J. What FERC should do to enable interregional electric transmission. *Electr. J.* **2024**, *37*, 107446. [\[CrossRef\]](#)
10. Yang, J.; Zhou, K. The impact of trans-regional power transmission on urban energy system resilience in China. *Energy* **2025**, *322*, 135555. [\[CrossRef\]](#)
11. Chen, Y.; Wang, H.; Yan, Z.; Xu, X.; Zeng, D.; Ma, B. A two-phase market clearing framework for inter-provincial electricity trading in Chinese power grids. *Sustain. Cities Soc.* **2022**, *85*, 104057. [\[CrossRef\]](#)
12. Mahata, S.; Harsh, P.; Shekher, V. Comparative study of time-series forecasting models for wind power generation in Gujarat, India. *e-Prime Adv. Electr. Eng. Electron. Energy* **2024**, *8*, 100511. [\[CrossRef\]](#)
13. Li, Y.F.; Zhang, Y.; Lin, F.; Zhao, Y.; Chen, Y.; Zhao, H.; Huo, W. Medium- and long-term power generation forecast based on climate characterisation and an improved XGBoost algorithm for photovoltaic power plants. *Power Syst. Prot. Control* **2024**, *52*, 84–92. [\[CrossRef\]](#)
14. Bai, R.; Li, J.; Liu, J.; Shi, Y.; He, S.; Wei, W. Day-ahead photovoltaic power generation forecasting with the HWGC-WPD-LSTM hybrid model assisted by wavelet packet decomposition and improved similar day method. *Eng. Sci. Technol. Int. J.* **2025**, *61*, 101889. [\[CrossRef\]](#)
15. Zhu, H.; Wang, Y.; Wu, J.; Zhang, X. A regional distributed photovoltaic power generation forecasting method based on grid division and TCN-BiLSTM. *Renew. Energy* **2026**, *256*, 123935. [\[CrossRef\]](#)
16. Gao, T.; Li, J.; Dai, Y.; Zhang, J.; Zhang, L.; Gao, N.; Hao, J.; Gao, W. Short-Term Load Forecasting of Multi-Energy in Integrated Energy System Based on Efficient Information Extracting Informer. *IET Renew. Power Gener.* **2026**, *20*, e70168. [\[CrossRef\]](#)
17. Wang, X.; Hu, M.; Luo, X.; Guan, X. Spatio-temporal photovoltaic power forecasting via Bayesian-optimized dynamic graph convolutional networks with temporal convolutional networks. *J. Clean. Prod.* **2026**, *542*, 147641. [\[CrossRef\]](#)
18. Piotrowski, P. Short-term forecasting of the total power generation from wind farms and solar power plants in the national power system using advanced ensemble machine learning models. *Energies* **2026**, *19*, 930. [\[CrossRef\]](#)
19. Regmi, S.; Rayamajhi, A.; Poudyal, R.; Adhikari, S. Evaluating Preparedness and Overcoming Challenges in Electricity Trading: An In-Depth Analysis Using the Analytic Hierarchy Process and a Case Study Exploration. *Electricity* **2024**, *5*, 271–294. [\[CrossRef\]](#)
20. Khayat, A.; Kissaoui, M.; Bahatti, L.; Raihani, A.; Atifi, Y.; Errakkas, K. A novel hybrid GRU-XGBoost model for day-ahead photovoltaic generation forecasting in microgrids. *Sci. Afr.* **2025**, *29*, e02884. [\[CrossRef\]](#)
21. Dai, Y.; Wang, Y.; Chen, Y.; Wu, J.; Chao, J. Combining meteorological and power information of station measurement and model prediction with the hybrid CNN-Transformer and CNN-BiLSTM for ultra-short-term photovoltaic power forecasting. *Int. J. Electr. Power Energy Syst.* **2025**, *171*, 111009. [\[CrossRef\]](#)
22. Mathew, A.; Chikte, R.; Sadanandan, K.S.; Abdelaziz, S.; Ijaz, S.; Ghaoud, T. Medium-term feeder load forecasting and boosting peak accuracy prediction using the PWP-XGBoost model. *Electr. Power Syst. Res.* **2024**, *237*, 111051. [\[CrossRef\]](#)
23. Zhu, C.; Tu, Y.; Wei, Q.; Zang, Y.; Zhou, P.; Yang, L.; Wang, J.; Yu, Y.; Lv, R.; Du, J.; et al. Short-term power prediction of photovoltaic power station based on LSTM-XGBoost model. *Sol. Energy* **2025**, *300*, 113819. [\[CrossRef\]](#)
24. Zhang, R.; Bu, S.; Zheng, Y.; Li, G.; Wan, X.; Zeng, Q.; Zhou, M. A novel multi-task learning model based on Transformer-LSTM for wind power forecasting. *Int. J. Electr. Power Energy Syst.* **2025**, *169*, 110732. [\[CrossRef\]](#)
25. Mohammed, R.H.; El-Saieed, A.M. Chaotic billiards optimized hybrid transformer and XGBoost model for robust and sustainable time series forecasting. *Sci. Rep.* **2025**, *15*, 25962. [\[CrossRef\]](#)

26. Mazen, F.M.A.; Shaker, Y.; Abul Seoud, R.A. Forecasting of solar power using GRU–Temporal Fusion Transformer model and DILATE loss function. *Energies* **2023**, *16*, 8105. [\[CrossRef\]](#)
27. Baidya, R.; Lee, S.W. Addressing the non-stationarity and complexity of time series data for long-term forecasts. *Appl. Sci.* **2024**, *14*, 4436. [\[CrossRef\]](#)
28. Ge, D.; Liu, B.; Cheng, Y.; Zhu, Z. Rethinking decomposition for time series forecasting: Learning from temporal patterns. *Expert Syst. Appl.* **2025**, *287*, 128044. [\[CrossRef\]](#)
29. Zhang, H.; Lin, J.; Zhang, W.; Yuan, H.; Li, G. AdaMixT: Adaptive weighted mixture of multi-scale expert Transformers for time series forecasting. In Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25), Montreal, QC, Canada, 16–22 August 2025; pp. 3633–3641. [\[CrossRef\]](#)
30. Du, L.; Gao, R.; Suganthan, P.N.; Wang, D.Z.W. Bayesian optimization-based dynamic ensemble for time series forecasting. *Inf. Sci.* **2022**, *591*, 155–175. [\[CrossRef\]](#)
31. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.
32. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [\[CrossRef\]](#)
33. Angelopoulos, A.N.; Bates, S. Conformal prediction: A gentle introduction. *Found. Trends Mach. Learn.* **2023**, *16*, 494–591. [\[CrossRef\]](#)
34. Jiang, Y.; Zhu, C.; Zou, F.; Zhang, L.; Yu, X.; Pan, C.; Liao, S. Stepwise conformal prediction for multi-step net load forecasting in microgrids under renewable energy variability. *Energies* **2026**, *19*, 2297. [\[CrossRef\]](#)
35. Meisenbacher, S.; Turowski, M.; Phipps, K.; Rätz, M.; Müller, D.; Hagenmeyer, V.; Mikut, R. Review of automated time series forecasting pipelines. *WIREs Data Min. Knowl. Discov.* **2022**, *12*, e1475. [\[CrossRef\]](#)
36. Alsharef, A.; Aggarwal, K.; Sonia; Kumar, M.; Mishra, A. Review of ML and AutoML solutions to forecast time-series data. *Arch. Comput. Methods Eng.* **2022**, *29*, 5297–5311, Correction in *Arch. Comput. Methods Eng.* **2023**, *30*, 3473. [\[CrossRef\]](#)
37. Zhu, Z.; Chen, W.; Xia, R.; Zhou, T.; Niu, P.; Peng, B.; Wang, W.; Liu, H.; Ma, Z.; Gu, X.; et al. Energy forecasting with robust, flexible, and explainable machine learning algorithms. *AI Mag.* **2023**, *44*, 377–393. [\[CrossRef\]](#)
38. de O. Santos, D.S., Jr.; de Mattos Neto, P.S.G.; de Oliveira, J.F.L.; Siqueira, H.V.; Barchi, T.M.; Lima, A.R.; Madeiro, F.; Dantas, D.A.P.; Converti, A.; Pereira, A.C.; et al. Solar irradiance forecasting using dynamic ensemble selection. *Appl. Sci.* **2022**, *12*, 3510. [\[CrossRef\]](#)
39. de Oliveira e Lucas, P.; Guimarães, F.G.; Mendes, E.M.A.M. Automated machine learning based on decomposition, causality and evolutionary multitask optimization for time series forecasting. *Neurocomputing* **2025**, *656*, 131569. [\[CrossRef\]](#)
40. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [\[CrossRef\]](#)
41. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
42. Amatriain, X.; Sankar, A.; Bing, J.; Bodigutla, P.K.; Hazen, T.J.; Kazi, M. Transformer models: An introduction and catalog. *arXiv* **2023**, arXiv:2302.07730. [\[CrossRef\]](#)
43. Wang, Y.; Wu, H.; Dong, J.; Qin, G.; Zhang, H.; Liu, Y.; Qiu, Y.; Wang, J.; Long, M. TimeXer: Empowering Transformers for Time Series Forecasting with Exogenous Variables. In Proceedings of the Advances in Neural Information Processing Systems 37; Neural Information Processing Systems Foundation, Inc. (NeurIPS): La Jolla, CA, USA, 2024. [\[CrossRef\]](#)
44. Zhao, H.; Xu, P.; Gao, T.; Zhang, J.J.; Xu, J.; Gao, D.W. CPTCFS: CausalPatchTST incorporated causal feature selection model for short-term wind power forecasting of newly built wind farms. *Int. J. Electr. Power Energy Syst.* **2024**, *160*, 110059. [\[CrossRef\]](#)
45. Open-Meteo. Previous Model Runs API. Available online: <https://open-meteo.com/en/docs/previous-runs-api> (accessed on 2 September 2026).
46. RTE. Données éCO2mix Nationales Consolidées et Définitives. Open Data Réseaux Énergies. Available online: <https://opendata.reseaux-energies.fr/explore/dataset/eco2mix-national-cons-def/> (accessed on 27 July 2026).
47. RTE; ODRÉ. Parc National Annuel de Production par Filière. Open Data Réseaux Énergies. Available online: <https://opendata.reseaux-energies.fr/explore/dataset/parc-prod-par-filiere/> (accessed on 29 July 2026).
48. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.