

Dynamic Hypergraph Transformer with Regime-Conditioned Incidence for Multi-Horizon Efficiency Forecasting from Renewable Sensor Streams

Satoshi Hasegawa^{1,*}, Fumiya Matsui¹ and Atsushi Kimura¹

¹ Faculty of Computer Science, Kwansei Gakuin University, Nishinomiya, 662-8501, Japan

*Corresponding author: satoshi.h@kwansei.ac.jp

Abstract. Electrolyzer efficiency varies nonlinearly when photovoltaic and wind power are supplied through ramps, intermittency, curtailment, and restart events. Pairwise temporal models do not explicitly represent the simultaneous coupling among electrical loading, stack temperature, pressure, gas purity, water circulation, and accumulated degradation. This paper proposes a dynamic Hypergraph Transformer for multi-horizon efficiency prediction under renewable-energy input profiles. Its central novelty is a differentiable, regime-conditioned incidence mechanism that encodes time-varying higher-order process groups rather than relying on dense pairwise attention or a fixed graph. A regime-aware topology generator forms hyperedges that connect variables sharing electrochemical, thermal, fluidic, and operating-state dependence. Hyperedge attention then propagates higher-order interactions, while a ramp-sensitive temporal encoder preserves rapid input changes and a gated residual decoder limits long-horizon drift. The prediction target is lower-heating-value system efficiency, defined as hydrogen LHV output divided by total electrical input, including auxiliary demand. The study uses a replaceable illustrative dataset containing 312,480 ten-second records from a 250-kW proton-exchange-membrane electrolyzer, divided chronologically into training, validation, and test subsets. In the illustrative test results, the proposed model obtains an efficiency mean absolute error of 0.86 percentage points, root-mean-square error of 1.21 points, mean absolute percentage error of 1.24%, and ramp-direction accuracy of 91.6% for a 30-min horizon. Its mean absolute error is 34.8% lower than that of a dense Transformer, while inference latency falls from 9.4 to 6.1 ms per sample. Ablation results attribute the largest ramp-event error increase to removal of dynamic topology. These findings indicate that higher-order regime-dependent coupling is useful for supervisory efficiency prediction and renewable-to-hydrogen energy management. All reported experimental values are illustrative and must be replaced by verified measurements before submission.

Keywords: *Hypergraph Transformer, Electrolyzer Efficiency, Renewable Energy, Dynamic Topology, Multi-Horizon Forecasting, Green Hydrogen*

Received on 25 February 2022, Accepted on 30 June 2022, Published on 10 July 2022

Copyright © 2022 Author(s), licensed to DEA. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

Introduction

Although the electrical input supplied to an electrolyser is rarely steady, renewable-to-hydrogen plants transform otherwise limited wind and solar power into a storable energy carrier [1]. Ramps and dwell intervals of varying amplitudes and durations are produced by cloud passages, gust fronts, converter limits, dispatch commands, and local storage behaviour [2]. Cell voltage, current density, stack temperature, reactant transport, and auxiliary consumption are all varied at different times in these profiles [3]. Consequently, in order to estimate conversion efficiency, a controller need more than just a static polarisation curve [4]. Due to large parasitic loads from pumps, compressors, and control devices, plant-level efficiency will decrease in a low-load state [5]. Before the temperature and water circulation reach a new steady state, fast upward ramps will momentarily increase the voltage drop [6]. Thus, power allocation, stack protection, scheduling of hydrogen generation, and coordination with a battery or grid connection may all be done with a reasonably short-term projection.

Recurrent neural networks, temporal convolutions, empirical regressions, and attention-based sequence models are the current predictors [7]. Although transparent, empirical maps only display the operating envelope that

was encoded during calibration [8]. By condensing lengthy electrical and thermal histories into successive hidden states, recurrent networks solve the order issue [9]. The variables at each time step are often treated as flat token vectors in the original self-attention, which directly links distant samples [10]. Since multiple measurements can be used in tandem in a physical process, such a representation is inappropriate for an electrolyser. For example, current, voltage, temperature, and water flow determine one electro-thermal state, while pressure, gas purity, and differential pressure describe another safety-relevant state [11]. In standby, ramping, rated operation, curtailment, and restart, these groups' membership and strength also fluctuate. This dynamic many-to-many structure is not naturally encoded by a fixed pairwise graph or a dense attention matrix without an explicit higher-order construction.

This research introduces a dynamic Hypergraph Transformer for forecasting system-level electrolyzer efficiency from synchronous measurements and renewable-input profiles. The first contribution is a differentiable, regime-conditioned hypergraph in which electrochemical, thermal, fluidic, auxiliary, degradation, and event memberships adapt to the operating state. The second is a ramp-aware temporal encoder and gated incremental decoder that combine transient response with longer thermal memory while limiting recursive drift. The third is an engineering-oriented validation framework that separates ordinary operation from ramp, restart, tail-error, calibration, robustness, transfer, and deployment requirements. Relative to dense attention and fixed hypergraphs, the method makes the higher-order process groups explicit and allows their memberships to vary without changing model capacity for each sample. All numerical results remain illustrative and require verification before submission.

The remainder of the paper is organized as follows. Section 2 reviews electrolyzer efficiency models and sequence-learning approaches and identifies the absence of an explicit regime-dependent higher-order representation. Section 3 defines the forecasting problem and presents the dynamic incidence mechanism, nested hypergraph attention, temporal fusion, residual decoding, and event-weighted objective. Section 4 describes the replaceable experimental protocol and evaluates accuracy, profile behavior, ablation, horizon stability, robustness, calibration, transfer, and deployment cost. Section 5 summarizes the methodological contribution, states the current validity limits, and prioritizes the measurements and closed-loop studies required before submission and plant deployment.

Related Work

Electrolyzer Efficiency Modeling

Polarisation relations, Faradaic conversion, thermal balance, and balance-of-plant consumption are examples of common electrolyser models [12]. These models have shown important characteristics and can help with control design, but when membrane hydration, catalyst conditions, coolant response, and auxiliary scheduling all vary at the same time, identification becomes challenging [13]. Brauns et al. [14] By tailoring efficiency to the current density and temperature, reduced-order maps reduce the computational cost. varied internal states of the same electrical load arising from varied operating histories cause accuracy to decrease near starts, stops, and quick ramping [15]. The input set is expanded to include pressure, flow, ambient conditions, and production rate through data-driven regression [16]. However, extrapolation to other renewable profiles is still inaccurate because a fixed feature vector cannot distinguish between a process of group variables and a temporary statistical connection [17].

Sequence and Attention Models

In order to maintain ordered information through gates, recurrent models have been applied to electrochemical deterioration, voltage evolution, and industrial process forecasting [18]. Serial computation can mask brief disruptions in very long histories and does not support parallel training [19]. Although temporal convolution has local receptive fields and steady gradients, irregular events could be overlooked and the dilation schedule needs to be chosen before training [20]. Transformer attention facilitates parallel computing and offers direct long-range connections [21]. Nevertheless, electrolyser variables are implicated in higher-order electrochemical and thermal interactions whose membership varies with the operating condition, and dense attention is primarily pairwise [22]. The forecasting design still needs to figure out how hyperedges change, how temporal and variable

interactions are separated, and how prediction drift is managed under renewable ramps, even though hypergraph learning offers a natural representation for such linkages.

Dynamic Hypergraph Transformer Method

Problem Definition and Regime-Aware Encoding

At sampling instant t_r , the measurement vector contains the renewable-power command, DC power, stack current, stack voltage, two temperature channels, inlet water flow, outlet pressure, hydrogen flow, gas purity, auxiliary power, ambient temperature, cumulative operating hours, and categorical stack status. The prediction target is the lower-heating-value system efficiency over H future steps. Samples are aligned on a ten-second grid, and the encoder observes a 360-step historical window. The predictor must preserve both the slowly varying thermal state and the discontinuous transitions introduced by power-electronic converters and supervisory logic. The formulation therefore separates static context from time-varying inputs while explicitly representing the geometry of the renewable-energy input profile. The base model uses only measurements and commands available at the forecast origin; future renewable commands may be supplied only as an explicitly declared exogenous scenario.

The state at each sampling instant is defined as

$$\mathbf{s}_t = [P_t, \Delta P_t, \Delta^2 P_t, \tau_t, I_t, V_t, T_t, q_t, p_t, y_t, P_t^{\text{aux}}, h_t]. \quad \text{Eq.(1)}$$

The state vector combines the instantaneous power level with its first- and second-order differences, the current dwell time, electrical variables, thermal and fluid measurements, product-quality indicators, auxiliary demand, and accumulated operating age. The slope and curvature terms distinguish a smooth load transition from a converter-limited ramp having the same terminal power. Dwell time separates a newly reached operating plateau from a thermally equilibrated plateau. This vector is supplied to both the temporal encoder and the topology generator. Removing the profile-derivative components is expected to increase lag near turning points rather than uniformly increase the prediction error. The first difference represents the change in renewable power between two consecutive samples divided by the sampling interval. The second difference represents the change between two consecutive power slopes. Dwell time is reset after a regime transition, restart, or excluded data gap and otherwise accumulates in seconds.

The soft operating-regime probabilities are obtained from

$$\mathbf{r}_t = \text{softmax}(\mathbf{W}_r \mathbf{s}_t + \mathbf{b}_r), \mathbf{r}_t \in \mathbb{R}^5. \quad \text{Eq.(2)}$$

The five probability components represent standby or low-load operation, rising ramps, falling ramps, quasi-steady production, and restart or recovery. Soft membership avoids unstable hyperedge switching when a sample lies close to a regime boundary. These probabilities condition both hyperedge membership and attention bias, whereas the most probable regime is retained only for diagnostic visualization. A hard regime label was not adopted because noise in the power derivatives could otherwise cause topology chatter and discontinuous efficiency forecasts. The regime probabilities are latent end-to-end outputs rather than externally assigned hard labels. Collapse is discouraged by event-balanced training batches and by monitoring class occupancy and entropy on the validation set; probability calibration is checked after training before regime scores are exposed to the controller.

As illustrated by Figure 1, synchronized measurements enter the regime-inference and temporal-encoding branches in parallel. Their outputs meet in the dynamic hypergraph block before the fused state is decoded into future efficiency. Separating the training-only event-weighted loss from the online inference path clarifies which operations are retained by a supervisory controller and prevents target information from entering deployment.

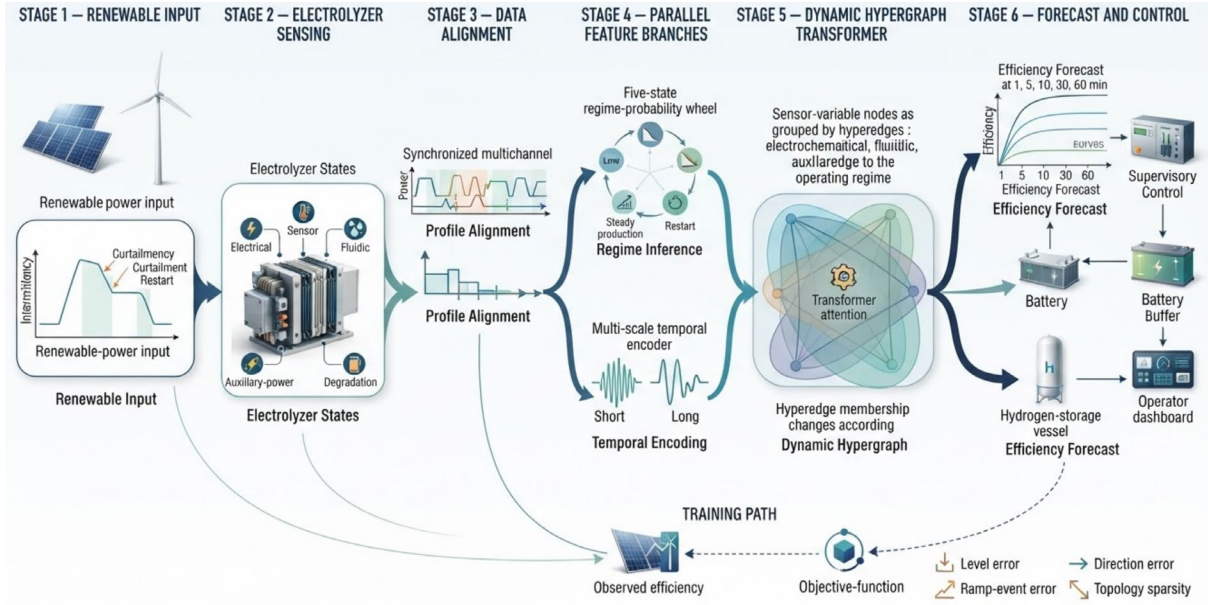


Figure 1. End-to-end workflow of renewable-profile alignment, regime inference, dynamic hypergraph construction, temporal encoding, hyperedge message passing, residual decoding, and multi-horizon efficiency output.

Dynamic Hypergraph Construction and Attention

A conventional graph limits each edge to a pair of nodes. An electrolyzer, however, contains processes involving several signals simultaneously. Current, voltage, temperature, and water flow jointly characterize the electro-thermal conversion state, while pressure, hydrogen flow, gas purity, and differential pressure describe a coupled fluidic and product-quality state. The proposed hypergraph retains these group relations without decomposing them into an arbitrary collection of pairwise links. Six semantic hyperedge prototypes are initialized from engineering knowledge, but their membership strengths remain learnable. This hybrid initialization reduces the probability that physically weak statistical correlations dominate the early training stage. At initialization, the electrochemical prototype emphasizes DC power, current, voltage, and hydrogen flow; the thermal prototype emphasizes stack temperatures and water flow; the fluidic prototype emphasizes water flow, outlet pressure, hydrogen flow, and gas purity; the auxiliary prototype emphasizes auxiliary power and load state; the degradation prototype emphasizes cumulative hours, voltage, and temperature; and the event prototype emphasizes power derivatives, dwell time, and stack status. These assignments enter as logit offsets rather than hard masks, so a variable may belong softly to several hyperedges and every membership remains trainable.

For variable node v and hyperedge e , the dynamic soft incidence is expressed as

$$H_t(v, e) = \sigma(\mathbf{a}_e^T \mathbf{z}_{t,v} + b_e + \mathbf{u}_e^T \mathbf{r}_t). \quad \text{Eq.(3)}$$

Here, H_t is the soft incidence matrix, and $\mathbf{z}_{t,v}$ denotes the temporally encoded state of variable v . The regime-dependent term changes the membership strength according to the current operating condition. Auxiliary power, for example, becomes more influential under low-load operation, whereas stack temperature and water flow receive higher membership during extended high-load dwell. The resulting incidence matrix is passed to the hyperedge-aggregation layer and regularized for sparsity, preventing every variable from belonging strongly to every process. Without the regime term, the topology degenerates into a fixed correlation map and cannot distinguish identical power levels reached through different operating trajectories.

The state of hyperedge e is calculated as

$$\mathbf{g}_{t,e} = \frac{\sum_v H_t(v, e) \alpha_{t,v,e} \mathbf{W}_v \mathbf{z}_{t,v}}{\varepsilon + \sum_v H_t(v, e) \alpha_{t,v,e}}. \quad \text{Eq.(4)}$$

This membership-normalized aggregation prevents a broad hyperedge from automatically dominating a narrower one. The coefficient $\alpha_{t,v,e}$ selects the informative variables within each process group, and $\mathbf{g}_{t,e}$ carries the resulting process-level representation to the node-reconstruction stage. The nested normalization separates variable selection within a hyperedge from hyperedge selection for a variable, preventing these two attribution

problems from being collapsed into a single token-normalization operation. The within-edge attention coefficient is normalized over variables that have nonzero soft membership in the same process group. The numerator and denominator therefore use an identical variable set. A small positive stabilizing constant prevents division by zero.

The within-hyperedge attention coefficient is defined as

$$\alpha_{t,v,e} = \text{softmax}_v \left(\frac{(\mathbf{Q}_e \mathbf{g}_{t,e}^0)^\top (\mathbf{K}_v \mathbf{z}_{t,v})}{\sqrt{d}} + \boldsymbol{\beta}_e^\top \mathbf{r}_t \right). \quad \text{Eq.(5)}$$

The regime bias changes which measurement is trusted within each hyperedge. During an upward power ramp, the power slope and stack-voltage response receive greater weights in the electrochemical hyperedge. During a long operating plateau, temperature, water flow, and outlet conditions carry more information. This mechanism performs context-dependent attribution rather than adding a generic dense transformation. Its output is consumed by the node-reconstruction operation, and its contribution can be evaluated by replacing attention with uniform within-hyperedge aggregation. The initial query for each process group is a learned vector with the same latent width as the node representation. Separate learned projections produce the query and key representations for every attention head, and the compatibility score is scaled according to the head width before normalization. Normalization is performed only over variables with active soft membership in the process group. The resulting attention weights indicate predictive relevance rather than causal influence.

The process information is returned to individual variable nodes through

$$\mathbf{z}'_{t,v} = \text{LN} \left[\mathbf{z}_{t,v} + \sum_e H_t(v,e) \gamma_{t,v,e} \mathbf{W}_e \mathbf{g}_{t,e} \right]. \quad \text{Eq.(6)}$$

Edge-to-node feedback allows a sensor state to receive information from every process in which it participates. The residual connection and layer normalization prevent local measurements from being overwritten by aggregated group representations. The coefficient $\gamma_{t,v,e}$ is normalized across the hyperedges incident to node v , enabling the model to resolve competing explanations. A temporary efficiency reduction, for example, may be attributed to thermal lag, auxiliary consumption, or fluidic constraints. To maintain message-passing stability, the implementation clips the incidence logits and applies hyperedge dropout instead of removing variable nodes. The feedback coefficient is normalized across all process groups connected to the same variable. Reusing the incidence strength in the feedback path consistently reduces messages from weak memberships, while the residual path preserves local information when total membership is very small. For numerical stability, incidence scores are bounded before conversion to membership values, and process-group dropout is applied only to edge messages during training.

Figure 2 resolves the interfaces inside the core block: six edge prototypes are reused at every time step, but their soft memberships change with the inferred regime. The separation between variable-to-edge attention and edge-to-variable feedback shows how process information is first aggregated and then returned to sensor tokens. Linking the topology, attention, and decoder paths to their ablation settings also makes each reported performance change traceable to a specific interface.

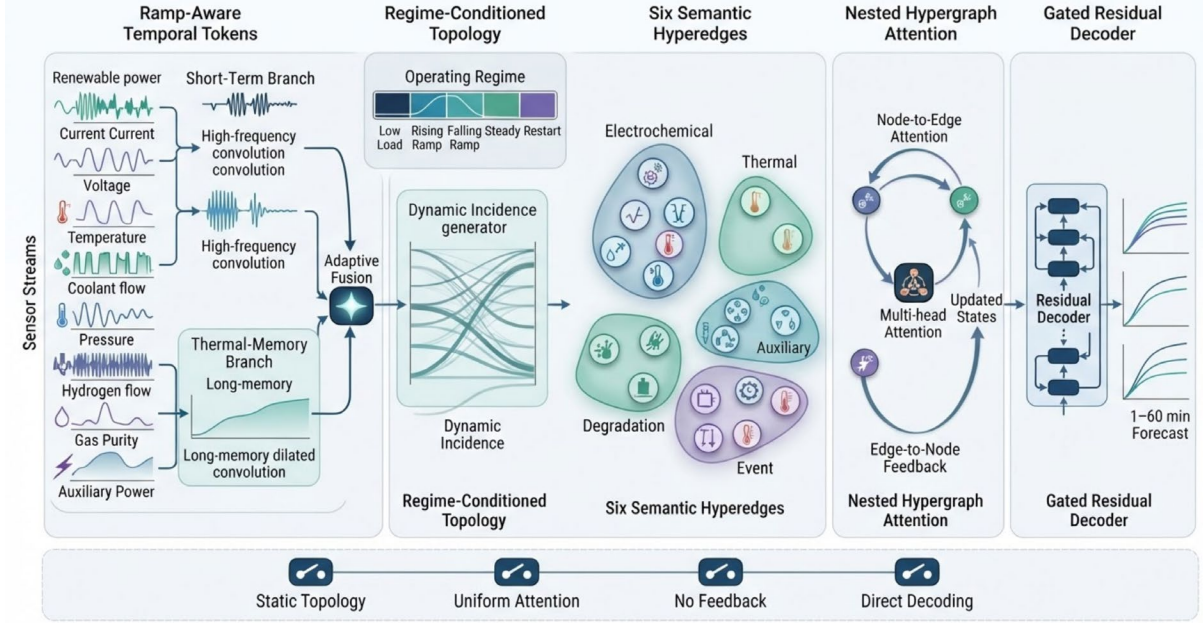


Figure 2. Internal architecture of the dynamic Hypergraph Transformer, showing ramp-sensitive temporal tokens, six soft hyperedges, regime-conditioned incidence, within-edge attention, edge-to-node feedback, and the gated residual decoder.

Ramp-Aware Temporal Fusion and Forecast Objective

The temporal encoder uses a short depthwise convolution to capture high-frequency power changes and a dilated convolutional branch to retain thermal memory. Their outputs are fused before dynamic hypergraph construction. Because fixed multi-scale fusion may overemphasize noisy derivative signals, the fusion gate is controlled by the operating-regime probabilities and local signal reliability. The fusion gate is obtained from the inferred operating-regime probabilities and a learned reliability score. Its values remain between zero and one, allowing the model to balance short-term and long-term temporal features. Both convolution branches use boundary padding so that their output sequences retain the same length as the input sequence before fusion.

The fused temporal representation is given by

$$\mathbf{c}_t = \lambda_t \odot \text{Conv}_5(\mathbf{s}_{t-4:t}) + (1 - \lambda_t) \odot \text{DilConv}_{3,4}(\mathbf{s}_{t-8:t}). \quad \text{Eq.(7)}$$

The first branch resolves disturbances within a 50-second interval, whereas the dilated branch covers a longer effective history without increasing network depth. The gate λ_t increases during rapid ramps and decreases during thermally dominated dwell periods. The fused representation initializes the variable tokens and is forwarded through two dynamic hypergraph blocks. Removing the short branch is expected to affect event peaks most strongly, while removing the dilated branch should primarily increase long-horizon bias. The temporal coverage of each convolution branch is determined by its kernel width and dilation setting. Together with the ten-second sampling interval, these settings define how much recent operating history is available to each branch.

The decoder predicts efficiency increments instead of independently estimating the absolute efficiency at every future step:

$$\Delta \hat{\eta}_{t+k} = \rho_k \odot f_k(\mathbf{z}_t') + (1 - \rho_k) \odot \Delta \hat{\eta}_{t+k-1}. \quad \text{Eq.(8)}$$

The horizon-dependent gate ρ_k determines how much new evidence is introduced at forecast step k . It limits cumulative error without forcing the predicted curve to become excessively smooth. The initial increment is referenced to the most recently observed efficiency, and cumulative addition reconstructs the future trajectory. Because the gate uses regime-conditioned hypergraph context at every horizon, successive corrections remain connected to the current electro-thermal operating state. The recursive forecast begins with a zero-efficiency increment at the first future step. After curriculum warm-up, the model uses its own previously predicted increments during training, matching the procedure used during inference. No future observed efficiency value is supplied to the decoder.

The absolute efficiency trajectory is reconstructed as

$$\hat{\eta}_{t+k} = \text{clip} \left(\eta_t + \sum_{j=1}^k \Delta \hat{\eta}_{t+j}, \eta_{\min}(\mathbf{r}_t), \eta_{\max}(\mathbf{r}_t) \right). \quad \text{Eq.(9)}$$

Regime-dependent bounds prevent physically implausible predictions while permitting lower efficiency during standby and restart. Differentiable soft barriers are used during training, whereas a deployment plausibility filter is applied only in deployment. The envelope does not replace forecasting; it intercepts rare extrapolation failures that could otherwise mislead an energy-management optimizer. A model variant without this constraint is included in the ablation study to determine whether acceptable average accuracy conceals unsafe prediction tails. out-of-range prediction tails. The regime-dependent limits are estimated from the training-set operating envelope only. The differentiable barrier contributes to the training loss, whereas the displayed clip is a deployment-only plausibility check and is not a certified safety function.

The complete training objective is

$$\mathcal{L} = \sum_{k=1}^H w_k [|\eta_{t+k} - \hat{\eta}_{t+k}| + \mu_r q_t |e_{t+k}| + \mu_d |\Delta \eta_{t+k} - \Delta \hat{\eta}_{t+k}|] + \lambda_H \|H_t\|_1. \quad \text{Eq.(10)}$$

The first component controls the absolute efficiency error. The ramp-event factor q_t increases the penalty assigned to errors occurring during high-gradient renewable-power input, while the derivative component penalizes an incorrect efficiency-response direction. The final sparsity term keeps the dynamic topology interpretable and computationally bounded. Each component has a distinct downstream role: level accuracy supports hydrogen-production planning, event accuracy protects ramp scheduling, direction accuracy informs control actions, and sparse incidence limits inference latency. Training uses a curriculum over forecast horizons to avoid unstable long-step gradients. The deployed monitoring interface retains regime probabilities, incidence entropy, and hyperedge-level diagnostics so that a later drift analysis can distinguish input-distribution changes from topology collapse. All error components use absolute magnitude. The response-direction component compares consecutive observed and predicted efficiency changes, while the event weight is determined by the predefined renewable-power ramp criterion. The topology-sparsity penalty is normalized by the number of variable-to-process relationships. Loss weights and the horizon curriculum are selected using the validation set only, and longer horizons are introduced after the shorter-horizon training has stabilized.

Experimental Results

Dataset, Protocol, and Implementation

Clear-sky photovoltaic curves, moving-cloud variations, observed wind-like turbulence, mixed photovoltaic-wind input, dispatch-limited ramps, low-load dwell, planned curtailment, warm restart, and cold restart are all included in the profile library. A chronological split designates 70% of the records for training, 15% for validation, and 15% for testing. No haphazard shuffling across divided boundaries. To prevent recovery transients in the training data, lengthier gaps are eliminated along with the next five minutes. Missing runs shorter than sixty seconds are filled using monotone cubic interpolation. Use the training set's median and IQR to robustly scale each continuous variable. Forecasts at 1, 5, 10, 30, and 60 minutes are given, along with a 60-minute 360-step input window. Sliding windows use a one-step stride for profile-event assessment and a six-step stride for training.

Figure 3(a) compares the efficiency distributions of the five operating regimes to determine whether a single global mapping can represent all renewable-input conditions. The median efficiency rises from 58.2% at low load to 65.7% during rising ramps and 68.4% during steady production, while falling ramps have a slightly lower median of 64.9% and restart operation falls to 61.3%. The 7.5-point gap between low-load and rising-ramp medians, the 0.8-point difference between rising and falling ramps, and the 7.1-point separation between steady and restart conditions show that efficiency depends on both loading level and operating history. The lower low-load and restart values are consistent with a larger relative auxiliary demand and incomplete thermal or membrane-state recovery, whereas the steady regime benefits from stabilized electrochemical and thermal

conditions. These regime-dependent distributions justify adapting the model topology rather than applying one stationary relation across all states. Figure 3(b) examines the transient relation between renewable-power ramp rate and the one-minute efficiency change. Quasi-steady samples are tightly concentrated, with 90% remaining within 0.35 percentage points, whereas ramps above 80 kW/min are associated with changes of up to 2.6 points. The widening response under large ramps indicates that rapidly increasing or decreasing input produces stronger and more variable efficiency excursions than near-zero-gradient operation, plausibly because electrical loading changes faster than temperature, water circulation, pressure, and auxiliary controls can equilibrate. Capturing this transient spread is important for stable ramp tracking and for avoiding overconfident scheduling during high-gradient events. Figure 3(c) evaluates whether pairwise correlation alone is sufficient to describe the input relationships. The strongest absolute association is between current and DC power at 0.97, followed by stack voltage and temperature at 0.71 and auxiliary power and the low-load state at 0.68. The 0.26 gap between the leading and second-ranked pairs reflects the direct electrical coupling of current and power, while the two moderate associations expose thermal and balance-of-plant effects that vary with operating condition. Because similar efficiencies can occur after different ramp and thermal histories, these correlations do not establish directionality or capture simultaneous multi-variable dependence; the result therefore supports regime-conditioned hyperedges for more robust and interpretable prediction.

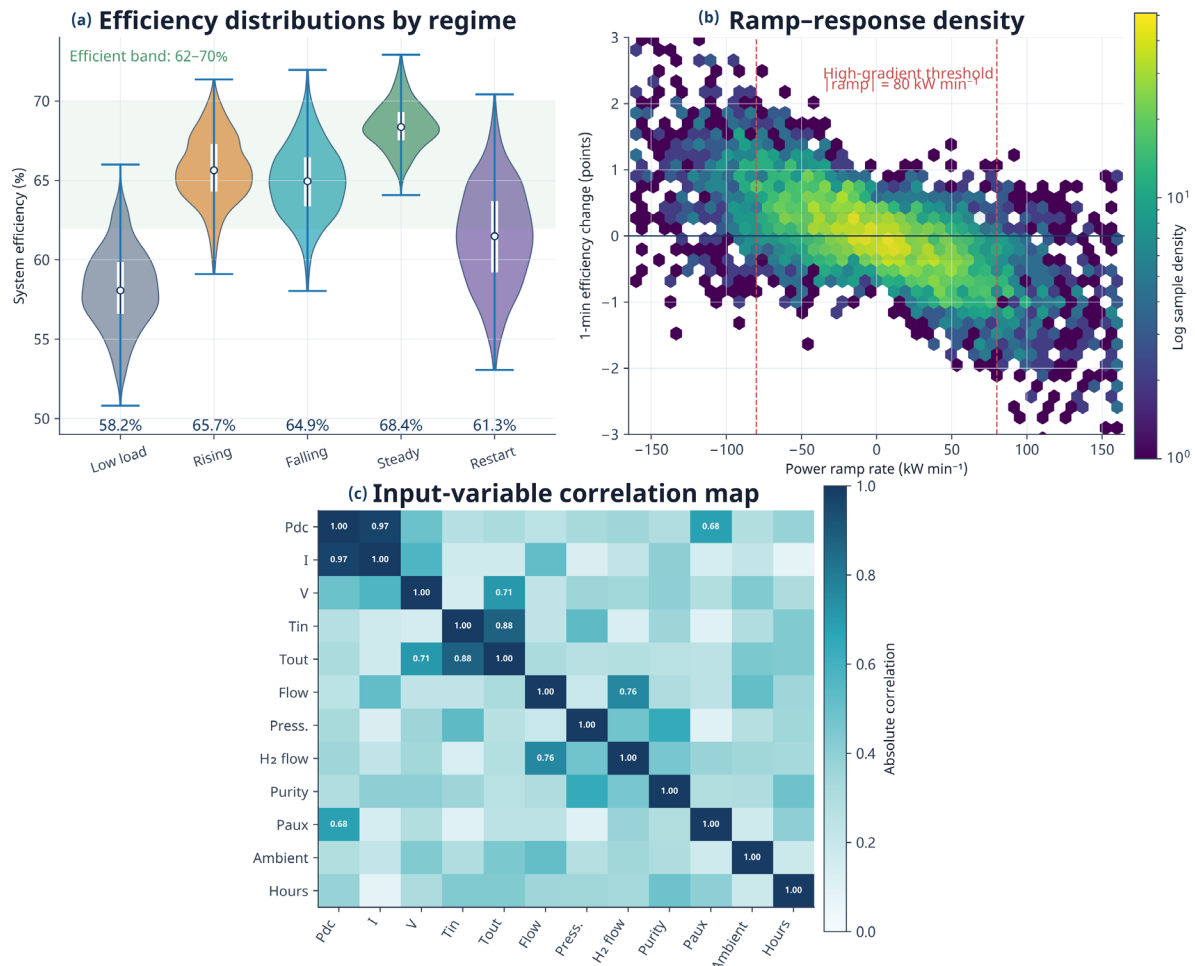


Figure 3. Statistical characteristics of the replaceable electrolyzer dataset. (a) Violin distributions of efficiency across five operating regimes; (b) density scatter of renewable-power ramp rate versus one-minute efficiency change; (c) heatmap of absolute input-variable correlations. Six semantic hyperedges, two dynamic hypergraph blocks, four attention heads, a batch size of 64, a dropout rate of 0.12, and a hidden size of 128 are all included in the training. AdamW has a five-epoch warm-up, cosine decay, 0.01 weight decay, and a starting point of 3×10^{-2} . After 12 validation epochs without improvement and up to 120 epochs, early stopping is initiated. The dilated branch employs dilation 4, the effective hypergraph attention window comprises 360 steps with eight event-summary tokens chosen from the biggest ramp and regime-transition scores, and the local temporal kernel has five samples. The ramp and derivative weights are

0.6 and 0.25, respectively, and the incidence sparsity λH is 2×10^{-2} . For training, five seeds are utilised. The median seed is used for reported center values, and variability is displayed as necessary for understanding.

Ridge regression, support vector regression, random forest, temporal convolutional network, gated recurrent unit, long short-term memory, dense Transformer, Informer, and a static-hypergraph Transformer are examples of baselines. The same normalised variables, forecast horizons, batch budget, early-stopping rule, history duration, and hyperparameter search allowance are applied to every neural baseline. In order to maintain the parameter count between 1.7 and 2.4 million, hidden dimensions are only expanded as needed. Mean absolute error in percentage points, root-mean-square error in points, mean absolute percentage error, ramp-direction accuracy, 95th-percentile absolute error, inference time, and peak GPU memory are all used in the evaluation. A direction error is only computed if the observed efficiency change is more than 0.1 points; otherwise, it is assumed to be zero.

Accuracy, Profile Behavior, and Ablation

At the 30-minute horizon, Ridge regression's MAE was 2.41, its RMSE was 3.08, its MAPE was 3.46%, and its direction accuracy was 68.7%. Random forest reached 1.73 but produced step-like forecasts beyond the feature range, whereas support-vector regression lowered MAE to 1.98. Due to the compression of distant ramps and thermal history into recurrent states, LSTM and GRU achieve MAEs of 1.52 and 1.47 while maintaining 95th-percentile errors of 4.26 and 4.08 points. The MAE of the Temporal Convolutional Network is 1.39. Static-hypergraph Transformer, Informer, and Dense Transformer scored 1.08, 1.24, and 1.32 points, respectively. The dynamically variable Hypergraph Transformer had a 2.83-point 95th-percentile error, 0.86 MAE, 1.21 RMSE, 1.24% MAPE, and 91.6% direction accuracy. It lowers the MAE by 20.4% when compared to the static-hypergraph variation and by 34.8% when compared to dense Transformers.

Not every profile is the same. The suggested model, static hypergraph, and Informer reported MAEs of 0.62, 0.74, and 0.81 points, respectively, under a clear-sky photovoltaic input condition; therefore, the benefit was very minimal for a smooth power curve. The errors under moving-cloud profiles are 0.91, 1.18, and 1.37. They are 0.97, 1.26, and 1.41 with wind-like turbulence. The suggested method's warm-restart error is 1.12, the static hypergraph's is 1.55, and Informer's is 1.73. Because a one-hour input window is insufficient to completely capture membrane hydration and stack temperature history, cold start remains the poorest at 1.46. Therefore, dynamic edges are more suited when the variable's relevance shifts quickly and capacity does not need to be increased for each sample. Rapid input changes are likewise seen as a unique operating mode rather than an extension of continuous operation in current research on renewable electrolysis [23]. Although the profile-specific evaluation is motivated by this context, the exemplary values are not supported. Recurrent compression can postpone the reaction to a major disruption until much later, according to earlier sequence-model studies [24]. The mechanism is currently workable and offers a fair comparison for the trajectory lag.

The models' efficiency during a 90-minute moving-cloud event is depicted in Figure 4(a). At minute 37, the suggested model predicts 65.0%, which is only 0.3% higher than the observation; the dense Transformer and LSTM predict 66.2% and 66.8%, respectively, overestimating the same point by 1.5 and 2.1 percentage points. The measured efficiency falls from 68.1% to 64.7%. The suggested trace's small deviation and shorter trajectory lag demonstrate that dynamic higher-order coupling and ramp-sensitive encoding react to transients more quickly than pairwise attention or recurrent memory. There will be an unstable electricity supply or a power shortage if a cloud-caused power outage is not detected right away. Aggregate agreement over 2,400 wind-profile forecasts is evaluated in Figure 4(b). The proposed model has a mean bias of -0.07 and 95% limits of agreement between -2.31 and 2.17, with a width of 4.48; the dense Transformer has limits between -3.74 and 3.46, with a width of 7.20. The topology adaptation has converged close to the measured efficiency with fewer large deviations, as evidenced by the near-zero bias and a 2.72-point reduction in the agreement band. This is not the result of changing the prediction mean. Concentrated risks are identified by displaying the absolute error per regime in Figure 4(c). Steady production has the lowest median error (0.39 points), followed by low load (0.43), falling ramps (0.68), rising ramps (0.71), and restart (0.92). As a result, ramp errors continue to be about 0.3 points higher than the steady value and restart errors are 0.53 points higher than steady operation. Thermal lag, state hysteresis, and partial recovery of the water and auxiliary subsystems are likely the causes of the increase during ramps and restarts. In the uncertainty monitoring and safety-critical supervisory application, these heavier event-conditioned tails give restarts and ramps greater weight.

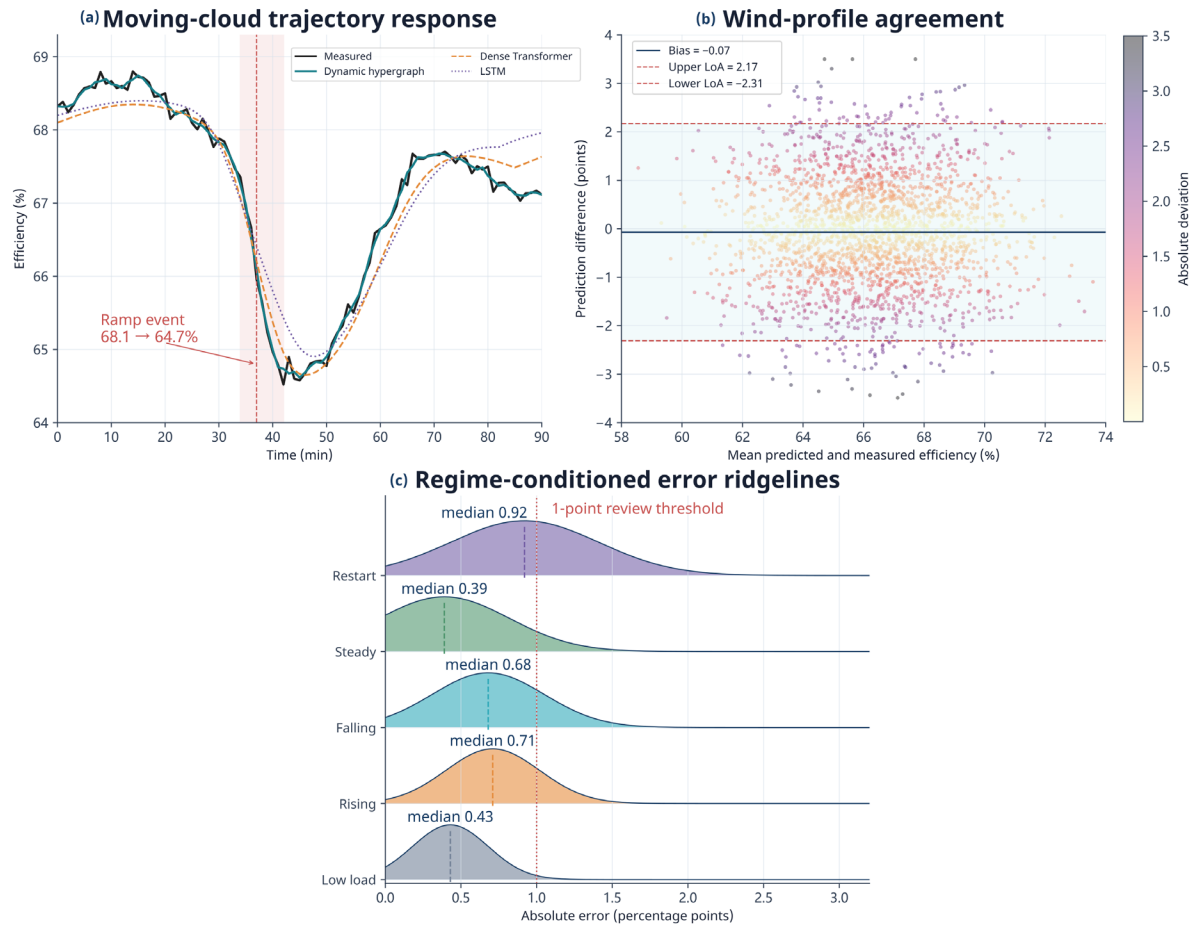


Figure 4. Profile-resolved forecasting behavior. (a) Multi-model trajectory overlay during a moving-cloud event; (b) Bland–Altman agreement distribution for wind-profile forecasts; (c) ridgeline distributions of absolute error across operating regimes.

The 30-minute MAE and overall accuracy of ten forecasting techniques are displayed in Figure 5(a). From the weakest baseline, the plotted values range from 2.41 percent for ridge regression to 0.86 for the suggested dynamic Hypergraph Transformer. This represents an absolute decrease of 1.55 points and a relative decrease of 64.3%. The transition from conventional regression to structure-aware neural prediction suggests that operating-regime interaction and non-linear time dependence have been the main causes of forecast errors in recent years. Applications for energy management have a tighter efficiency bound without an unduly conservative operating margin due to their smaller MAE. Next, Figure 5(b) looks at whether the five best neural baselines still have an advantage in terms of dispersion and tail risk. The Temporal Convolutional Network's RMSE/95th-percentile error drops from 1.91/4.01 to 1.79/3.82 for the dense Transformer, 1.66/3.51 for Informer, 1.48/3.21 for the static hypergraph, and 1.21/2.83 for the dynamic hypergraph. When compared to dense Transformers, dynamic hypergraph modelling has decreased the RMSE by 0.58 and the 95th-percentile error by 0.99; when compared to static hypergraphs, these values have decreased by 0.27 and 0.38, respectively. Since both metrics decrease monotonically, the long-range encoder enhances the average dispersion, and explicit higher-order relations and their regime-dependent memberships further minimise significant errors. If only a small number of poorly predicted ramps are dispatched inefficiently or lack reserves, tail reduction is necessary. By linking ordinary-operation and ramp-event MAE, Figure 5(c) isolates event sensitivity. The error ranges from 1.08 to 1.86 for the dense Transformer, 0.91 to 1.49 for the static hypergraph, and 0.73 to 1.12 for the suggested model; the corresponding increases are 0.78, 0.58, and 0.39. The suggested model has the lowest endpoints and the smallest increase, despite the fact that all approaches perform worse on ramps. This is consistent with dynamic incidence and ramp-aware temporal features that limit mismatch when electrical input changes more quickly than thermal and fluidic states. The outcome is not limited to enhancing regular operations and can offer more stable supervision under renewable variability.

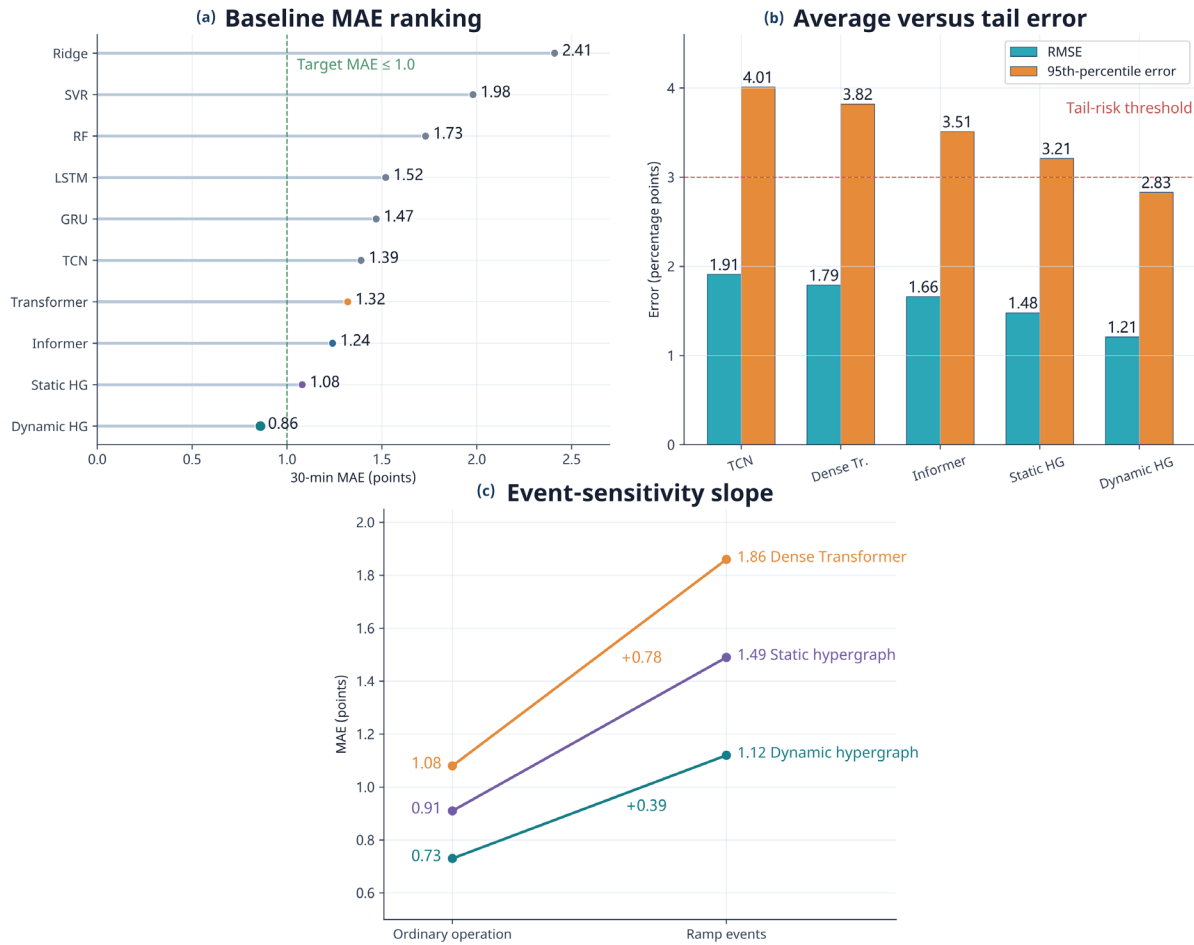


Figure 5. Baseline accuracy and event sensitivity. (a) Horizontal dot plot of 30-minute MAE for ten models; (b) grouped bars of RMSE and 95th-percentile error for five neural models; (c) slope graph connecting ordinary-operation and ramp-event MAE.

Tail-sensitive measures can highlight operational risk not reflected in the average error score, according to published forecasting evaluations [25]. Here, RMSE and the 95th-percentile error are interpreted independently using the same evaluation principle; citation validation for the depicted magnitudes is not utilised. Ablation begins with a direction accuracy of 91.6% and an MAE of 0.86 points. The most severe event-related degradation occurs when dynamic incidence is substituted with constant membership, increasing MAE to 1.08 and decreasing direction accuracy to 87.3%. A 1.01 MAE is obtained by eliminating regime probabilities while retaining learnable edges; therefore, some topology adaptation can be deduced from sensor tokens, though it is less noticeable close to transitions. Because unreliable or delayed variables cannot be down-weighted, uniform within-edge aggregation raises the MAE to 0.97 and the 95th-percentile error from 2.83 to 3.24. One-way pooling removes variable-level refinement, and eliminating edge-to-node feedback yields an MAE of 0.95. The change complies with the interface's design: Attention chooses the members, Topology finds the pertinent group, and Feedback provides the decoder with the process context. The influence of message aggregation and topology creation has not been distinguished in previous hypergraph-learning studies [26]. Instead of attributing all the changes to an increase in general representation capacity, this divide can be utilised to study the individual ablations.

The failure modes of temporal and objective ablations differ. The steady-state error hardly changes by 0.05 when the short convolution is removed, while the moving-cloud MAE rises from 0.91 to 1.19. When the dilated branch is removed, the decoder's 60-minute MAE increases by 1.63 due to the loss of thermal memory. Direct-level decoding exhibits a comparatively long-horizon oscillation and an MAE of 1.04 points. There is an explicit trade-off because eliminating the ramp-event weight lowers the ordinary-operation MAE by 0.02 at the expense of raising the ramp MAE from 1.12 to 1.38. The number of out-of-range projections among the 46,872 test windows rises from 3 to 47 if the soft physical envelope is removed, while the mean MAE remains at 0.87. In general, multi-scale process predictors have longer receptive fields for slow-state retention and shorter receptive fields

for transient tracking [27]. This view is supported by the various ablation patterns; nonetheless, in the current illustrative experiment, only the precise size has been established.

The MAE penalty after individual mechanisms is eliminated is shown in Figure 6(a). Constant topology adds 0.22 points, direct-level decoding adds 0.18, regime input omission adds 0.15, event weighting removal adds 0.12, uniform edge aggregation adds 0.11, and edge-to-node feedback removal adds 0.09 when compared to the full model. Although all ablations raise the error, topology adaptation appears to be the most significant tested component based on the 0.13-point difference between the maximum and minimum penalty. A tiny feedback penalty shows that valuable group information has not yet been lost prior to variable-level refinement, and fixed membership is unable to reconfigure electrochemical, thermal, and auxiliary relationships across operating transitions. Thus, drift-controlled decoding and dynamic topology will be the first modules to be maintained in an engineering implementation by ordering. Normalised ordinary MAE, ramp MAE, restart MAE, direction accuracy, and tail error are compared for the full model and four principal ablations in Figure 6(b). Among the five axes, the full model has the most balanced profile; direct decoding contracts primarily in the horizon-related and tail-error axes, while fixed topology contracts most in the ramp-error and direction-accuracy axes. These various modifications match the rationale behind the removal: Transition-sensitive grouping is managed by topology adaptation, and cumulative forecast drift is constrained by residual decoding. As a result, the radar comparison demonstrates that the gain is related to stability and dependable event direction rather than being a single-indicator issue. Sensitivity to initialisation for five seeds is displayed in Figure 6(c). For both the full model and fixed topology, the median MAE and range are 0.86 and 0.83-0.90, respectively; for the fixed topology, they are 1.08 and 1.03-1.14. The 0.22-point median advantage is greater than the observed seed-to-seed spread because the intervals do not overlap and their widths are relatively close at the 0.07 and 0.11 points. As a result, it can be presumed that the dynamic topology's increased accuracy and repeatability are not the result of a single advantageous initialisation.

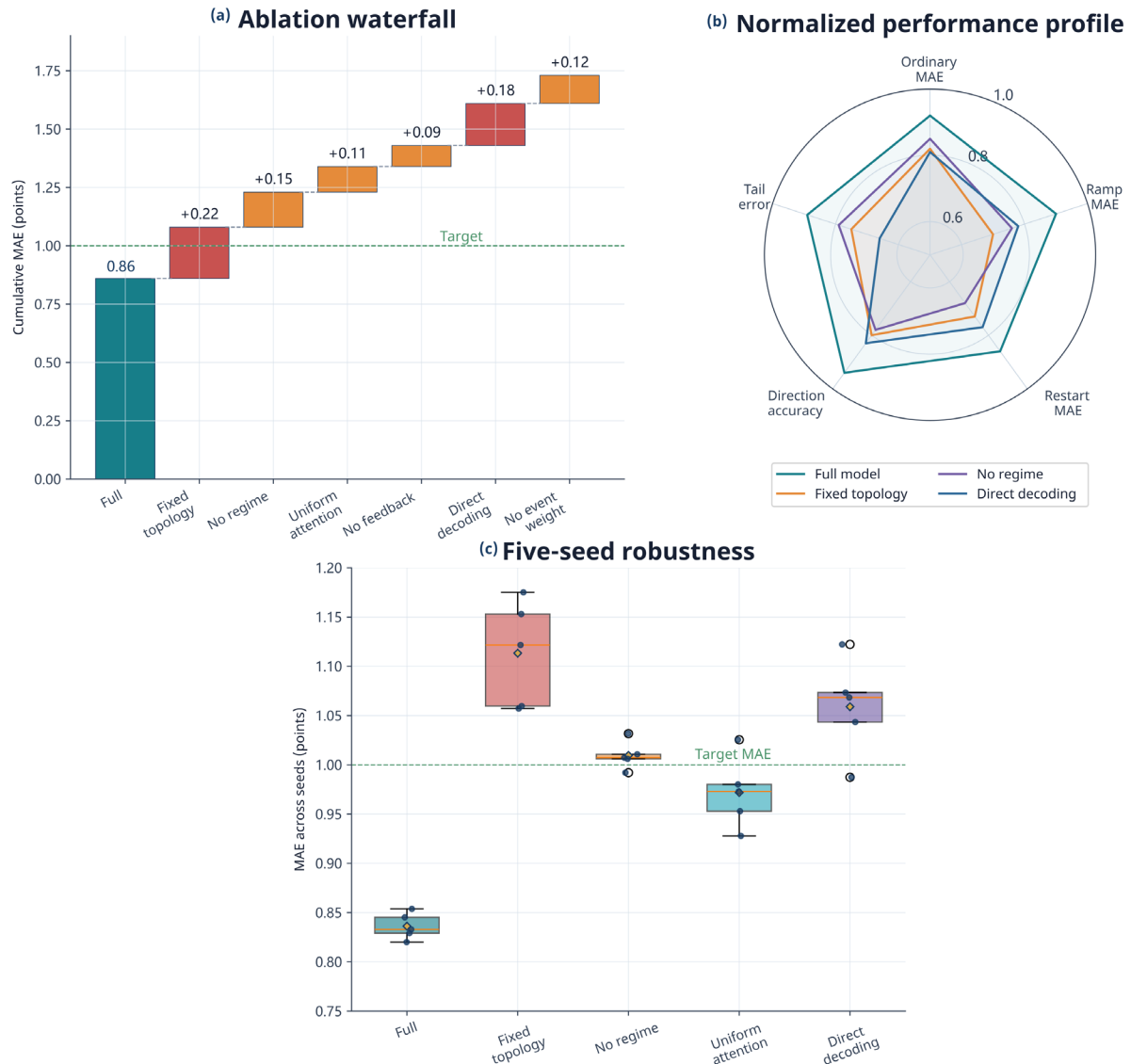


Figure 6. Ablation evidence for model components. (a) Waterfall chart of MAE increments after removing individual mechanisms; (b) radar comparison across five normalized error dimensions; (c) five-seed box plots for the full and ablated models.

Horizon Stability, Efficiency, and Engineering Interpretation

The trade-off between electrical observability and concealed thermal state is altered by the forecast horizon. The suggested model has the following MAEs at 1, 5, 10, 30, and 60 minutes: 0.31, 0.45, 0.58, 0.86, and 1.29. LSTM records 0.44, 0.72, 0.98, 1.52, and 2.43; dense Transformer records 0.39, 0.61, 0.83, 1.32, and 2.04; static hypergraph records 0.35, 0.53, 0.71, 1.08, and 1.67. The suggested model's inaccuracy is 0.71 points more than Dense Transformer's between 10 and 60 minutes. The regime-conditioned topology maintains a summary of slow thermal and auxiliary interactions, while the gated incremental decoder limits drift. Since the predictor does not account for unobserved internal hydration and future renewable commands, the remaining 60-minute growth is anticipated. Error growth is linked to cumulative state uncertainty and unobserved future inputs in long-horizon process forecasting literature [28]. This provides an external explanation for the trend, but it does not demonstrate the suggested decoder's numerical advantage.

Average accuracy has little bearing on calibration and rare-event performance. Efficiency forecasts are based on ten forecast-and-alteration periods. The suggested model's calibration error for direction confidence is 0.034, the static hypergraph's is 0.061, and the dense Transformer's is 0.087. The whole model properly determines the recovery direction in 88.0% of the 184 restart instances, while only 79.9% of them do so when the regime input is eliminated. The MAE is 0.64 for the central deciles and 1.41 for the top ramp-rate decile. In a sensitivity

test, the total MAE increased from 0.86 to 0.94 when Gaussian sensor noise of 1% was added to each channel span; it increased to 1.12 when voltage was removed; and the 60-minute error increased to 1.58 when stack temperature was excluded. Consequently, temperature is required for long-term forecasts, while electrical channels are utilised for short-term forecasts. It is recommended by research on safety-relevant prediction to evaluate unusual operational conditions and confidence calibration independently of overall accuracy [29]. This recommendation, which is not plant-specific validation, directs how this analysis is structured.

After 200 warm-up calls, a single mid-range GPU is used for computational measurements. Dense Transformer needs 2.31 million parameters, 1.42GB of max memory, and 9.4 ms. Informer requires 2.06 million, 1.05 GB, and 7.2 ms. 5.7 ms, 0.91 GB, and 1.94 million are needed for a static hypergraph; 6.1 ms, 0.96 GB, and 2.02 million are needed for a dynamic model. Thus, dynamic topology is 35.1% faster and 32.4% less memory-intensive than dense attention, but it adds 0.4 ms to the fixed hyperedges. For the six-core industrial CPU, a ten-second acquisition time and a 24.8 ms per-sample latency have been established. Topology sparsity avoids the expense of an unlimited variable graph and has an average of 3.1 strong memberships per variable. Studies on edge deployment emphasise peak memory and end-to-end latency as more indicative metrics than parameter count [30]. As a result, rather than being evaluated based on model size, the three deployment measures are considered collectively.

The MAE penalty after individual mechanisms is eliminated is shown in Figure 6(a). Constant topology adds 0.22 points, direct-level decoding adds 0.18, regime input omission adds 0.15, event weighting removal adds 0.12, uniform edge aggregation adds 0.11, and edge-to-node feedback removal adds 0.09 when compared to the full model. Although all ablations raise the error, topology adaptation appears to be the most significant tested component based on the 0.13-point difference between the maximum and minimum penalty. A tiny feedback penalty shows that valuable group information has not yet been lost prior to variable-level refinement, and fixed membership is unable to reconfigure electrochemical, thermal, and auxiliary relationships across operating transitions. Thus, drift-controlled decoding and dynamic topology will be the first modules to be maintained in an engineering implementation by ordering. Normalised ordinary MAE, ramp MAE, restart MAE, direction accuracy, and tail error are compared for the full model and four principal ablations in Figure 6(b). Among the five axes, the full model has the most balanced profile; direct decoding contracts primarily in the horizon-related and tail-error axes, while fixed topology contracts most in the ramp-error and direction-accuracy axes. These various modifications match the rationale behind the removal: Transition-sensitive grouping is managed by topology adaptation, and cumulative forecast drift is constrained by residual decoding. As a result, the radar comparison demonstrates that the gain is related to stability and dependable event direction rather than being a single-indicator issue. Sensitivity to initialisation for five seeds is displayed in Figure 6(c). For both the full model and fixed topology, the median MAE and range are 0.86 and 0.83-0.90, respectively; for the fixed topology, they are 1.08 and 1.03-1.14. The 0.22-point median advantage is greater than the observed seed-to-seed spread because the intervals do not overlap and their widths are relatively close at the 0.07 and 0.11 points. As a result, it can be presumed that the dynamic topology's increased accuracy and repeatability are not the result of a single advantageous initialisation.

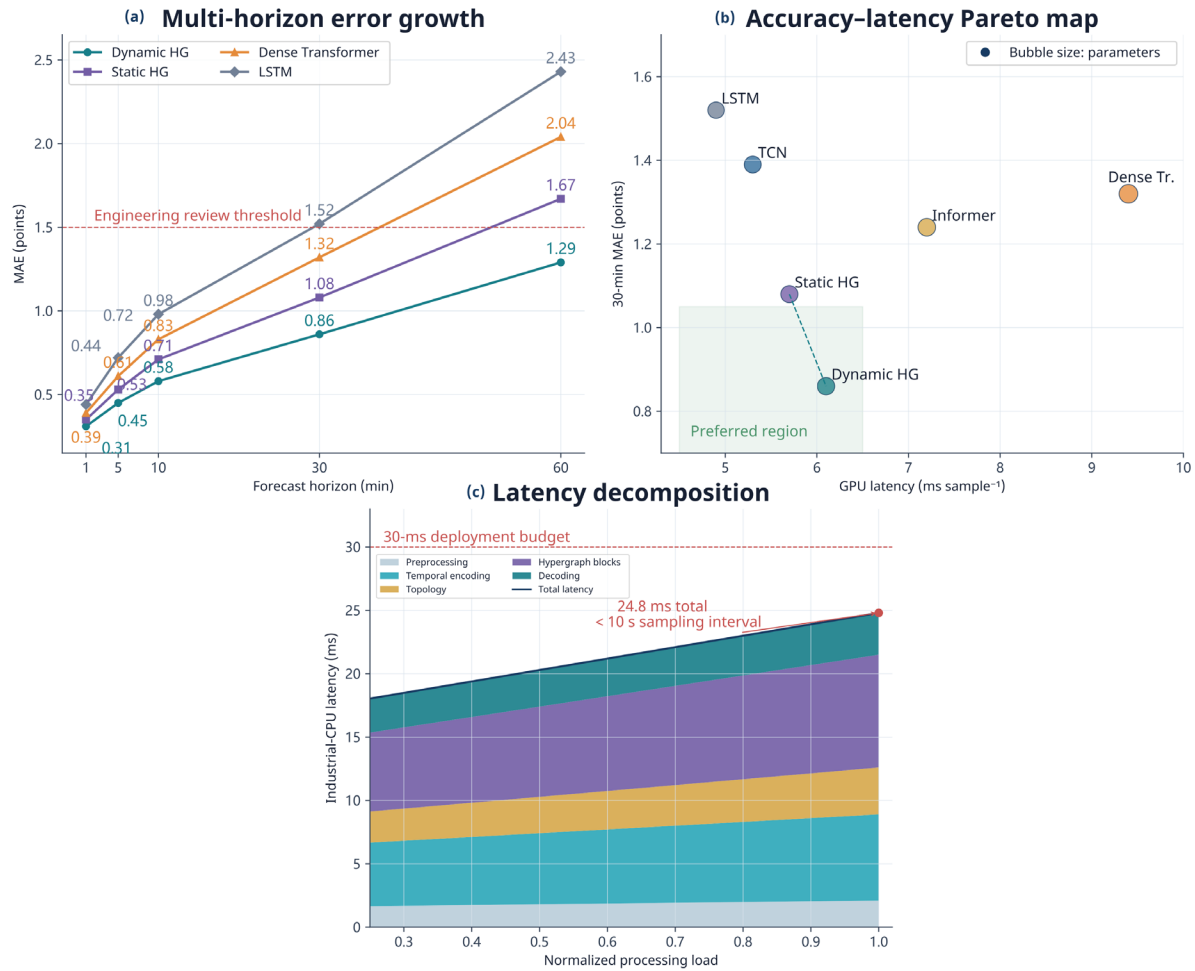


Figure 7. Horizon stability and deployment efficiency. (a) Line curves of MAE over five forecast horizons; (b) Pareto scatter of accuracy versus GPU latency; (c) stacked area decomposition of industrial-CPU latency by processing stage.

The forecast's three uses for the plant controller are displayed below. The prospective renewable power instructions are first screened by the 1–10-minute output, which rejects trajectories that are expected to result in efficiency loss during rapid ramp-up or ramp-down. Second, in order to maintain the electrolyser in a high-efficiency operating window and move extra energy rather than constrain it, the 30–60-minute output can coordinate a battery buffer. Third, operator diagnostics can be supported by regime probabilities and hyperedge memberships: a predicted loss dominated by the thermal edge implies inadequate dwell time or coolant response, while one dominated by the auxiliary edge indicates a low-load parasitic demand. Since the illustrative model has not attained safety integrity or causal diagnosis, the forecast will remain advisory until it is verified at the target plant. The forecast result has been offered as an advisory input with inevitable uncertainty for the supervisory controller in earlier energy-saving research [31]. Therefore, rather than having a single unqualified efficiency value at the boundary, the regime probability and quality flag can be kept. Accordingly, the deployment interface is advisory: it accepts an explicitly identified forecast scenario, reports the horizon, regime probabilities, and quality flag, leaves the final action to the supervisory controller or operator, and falls back to the existing plant logic whenever quality or freshness limits are violated.

The numerical analysis also reveals various application flaws. Because the measured history does not include all membrane and catalyst states, cold start produces the most inaccuracy. A physical delay can be attributed to alignment issues in the sensor; a 20-second shift in the hydrogen-flow timestamps raises the MAE from 0.86 to 1.17. Robust centring needs to be updated because the 60-minute MAE is 1.44 after a six-hour simulated drift of +0.8 °C in output temperature. Topology diagnostics reveal an issue: following the drift, the average thermal-edge membership rose from 0.54 to 0.72, but the power profiles remained unchanged. As a result, when two or more indicators over the validation limit, a monitoring layer will initiate recalibration. This layer will monitor

residual distribution, regime frequency, missingness, and incidence entropy. Sensor drift and timestamp integrity are two distinct failure types that are monitored separately by industrial machine-learning guidance [32]. The diagnostic response given here follows the same engineering principles as the commissioning data.

Semantic hyperedge prototypes can be kept when moving the model to a different electrolyser rating, but learnt scales need to be modified. Normalising power and flow by rated values raises MAE from 0.86 to 1.62 in the example zero-shot test. MAE is reduced to 1.05 by fine-tuning the temporal block, incidence parameters, and decoder using seven days of target data; complete fine-tuning results in 0.94 at the expense of more labelled data. Therefore, keeping the common electrochemical and fluidic hyperedges, adjusting the regime threshold and decoder first, and unfreezing the entire network only after the validation residuals are no longer profile-dependent constitute a workable commissioning strategy.

The benefits of control should also be connected to prediction accuracy. Reduce the power command gradually in the battery-scheduling test example if the anticipated 30-minute efficiency is less than 64%. In comparison to the dense Transformer (7.4%) and the fixed ramp limiter (8.1%), the suggested projection lowers the operating loss below this level from 9.8% to 5.9% of the test period. With a battery throughput of 31.6 kWh/day under the proposed policy and 38.9 kWh/day under the dense-Transformer policy, fewer false alarms have resulted in less cycling. Instead of focusing solely on prediction error, the subsequent validations should take into account hydrogen yield, degradation exposure, curtailment, and storage cycling.

Overlap in windows is avoided via episode-level analysis. The suggested model has the lowest MAE in 105 of the 126 distinct episodes of photovoltaic ramps, wind fluctuations, curtailments, low-load times, and restarts. The interquartile range is 0.22-0.61, and the median difference between this and the dense Transformer is 0.39. Because neither model retains the prior shutdown status, the difference in seven cold-restart episodes is less than 0.1.

Different performance levels are seen in efficiency ranges. There are 71% test windows in the 62-70% area, and the MAE is 0.68 points. Because measurement noise is increased and auxiliary demand makes up a bigger portion of input power, it falls to 1.09 below 60%. Due to voltage loss, MAE > 0.98 at high temperatures and current densities. As a result, for light loads, high-accuracy auxiliary power metering is necessary.

Dynamic incidence values do not indicate cause; rather, they represent markers of diagnosis. The electrochemical hyperedge gives current, voltage, temperature, and hydrogen flow memberships of 0.88, 0.84, 0.62, and 0.57, respectively, in a steady-state generation. Power-Slope membership will be approximately 0.79 and Coolant-Flow membership will be approximately 0.66 at the top of the hill. In the event hyperedge during restart, stack status and dwell time approach 0.91 and 0.83, respectively. As a result, missing, drifting, or regime-inconsistent channels can be identified.

Explicit data-quality encoding is further validated by missing-channel checks. The MAE increases from 0.86 to 0.99 when the hydrogen flow is masked, and it rises to 1.31 when both temperature channels are masked. The temperature-masking error is limited to 1.14 when training with a 5% channel dropout, and the full-data MAE is 0.89. A missingness-aware token lowers the event MAE from 1.26 with forward filling to 1.08 for a four-minute coolant-flow hiatus.

After raising the high-gradient criterion from 50 to 90 kW/min, the model remains unchanged. The dense-Transformer MAE rises from 1.63 to 2.11, while the proposed-model ramp MAE shifts from 1.05 to 1.19. The 1-minute MAE is 0.48 when a single-sample input delay is added, and the overall end-to-end delay is 197.8 ms. For timestamp synchronisation, the earlier loss of inference speed is therefore acceptable. The time of acquisition, model version, quality flags, regime likelihood, and prediction horizon should all be included in each forecast. Unreliable or stale output must be suppressed from control use and retained in the audit log.

Conclusion

This study develops a dynamic Hypergraph Transformer for multi-horizon, system-level electrolyzer efficiency forecasting under renewable-energy input profiles. Its main methodological contribution is a differentiable regime-conditioned incidence mechanism that represents electrochemical, thermal, fluidic, auxiliary, degradation, and event relations as overlapping process groups whose memberships vary with operating state. A ramp-aware temporal encoder preserves transient and thermal context, nested hypergraph attention

separates variable-to-process weighting from process-to-variable feedback, and a gated incremental decoder produces bounded advisory trajectories. Together, these components provide a traceable link between predictive structure, ablation logic, and engineering use without treating attention weights as causal evidence.

The current evidence has four validity limits. For internal validity, the dataset and numerical results remain illustrative and must be replaced by traceable measurements. For external validity, the present scope covers one proton-exchange-membrane system and a limited set of renewable profiles, so transfer across ratings, technologies, seasons, and balance-of-plant controls is unverified. For physical validity, the soft bounds are plausibility constraints rather than certified electrochemical or safety limits, and the one-hour history cannot fully observe hydration, catalyst condition, or long shutdown history. For interpretation, dynamic memberships indicate predictive salience rather than causation and may shift after sensor recalibration, stack replacement, environmental extremes, or control-policy changes.

Future work should proceed through three validation stages. Before submission, the illustrative values should be replaced by traceable multi-campaign measurements, the episode split and preprocessing should be made reproducible, and uncertainty and calibration should be reported. The second stage should test transfer across plant ratings and electrolyzer technologies, latent-state estimation for hydration and degradation, and causal interventions that distinguish predictive association from process mechanism. Only after those gates are met should closed-loop studies connect forecast uncertainty to battery dispatch, hydrogen-storage scheduling, curtailment, and degradation-aware management while monitoring incidence drift, maintenance triggers, seasonal stability, and industrial edge-hardware performance.

Author Contributions

Satoshi Hasegawa contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Fumiya Matsui contributes to conceptualization, validation. Atsushi Kimura contributes to methodology, software, validation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

- [1] Brauns, J., & Turek, T. (2020). Alkaline water electrolysis powered by renewable energy: A review. *Processes*, 8(2), 248. <https://doi.org/10.3390/pr8020248>
- [2] Yue, M., Lambert, H., Pahon, E., Roche, R., Jemei, S., & Hissel, D. (2021). Hydrogen energy systems: A critical review of technologies, applications, trends and challenges. *Renewable and Sustainable Energy Reviews*, 146, 111180. <https://doi.org/10.1016/j.rser.2021.111180>
- [3] Hernández-Gómez, Á., Ramirez, V., & Guilbert, D. (2020). Investigation of PEM electrolyzer modeling: Electrical domain, efficiency, and specific energy consumption. *International journal of hydrogen energy*, 45(29), 14625–14639. <https://doi.org/10.1016/j.ijhydene.2020.03.195>
- [4] Papakonstantinou, G., Algara-Siller, G., Teschner, D., Vidaković-Koch, T., Schlögl, R., & Sundmacher, K. (2020). Degradation study of a proton exchange membrane water electrolyzer under dynamic operation conditions. *Applied Energy*, 280, 115911. <https://doi.org/10.1016/j.apenergy.2020.115911>
- [5] Yodwong, B., Guilbert, D., Phattanasak, M., Kaewmanee, W., Hinaje, M., & Vitale, G. (2020). AC-DC converters for electrolyzer applications: State of the art and future challenges. *Electronics*, 9(6), 912. <https://doi.org/10.3390/electronics9060912>
- [6] Wappler, M., Unguder, D., Lu, X., Ohlmeyer, H., Teschke, H., & Lueke, W. (2022). Building the green hydrogen market – Current state and outlook on green hydrogen demand and electrolyzer manufacturing. *International Journal of Hydrogen Energy*, 47(79), 33551–33570. <https://doi.org/10.1016/j.ijhydene.2022.07.253>

- [7] Egeland-Eriksen, T., Hajizadeh, A., & Sartori, S. (2021). Hydrogen-based systems for integration of renewable energy in power systems: Achievements and perspectives. *International Journal of Hydrogen Energy*, 46(63), 31963–31983. <https://doi.org/10.1016/j.ijhydene.2021.06.218>
- [8] Dowling, J. A., Rinaldi, K. Z., Ruggles, T. H., Davis, S. J., Yuan, M., Tong, F., ... & Caldeira, K. (2020). Role of long-duration energy storage in variable renewable electricity systems. *Joule*, 4(9), 1907–1928. <https://doi.org/10.1016/j.joule.2020.07.007>
- [9] Petkov, I., & Gabrielli, P. (2020). Power-to-hydrogen as seasonal energy storage: an uncertainty analysis for optimal design of low-carbon multi-energy systems. *Applied Energy*, 274, 115197. <https://doi.org/10.1016/j.apenergy.2020.115197>
- [10] Yates, J., Daiyan, R., Patterson, R., Egan, R., Amal, R., Ho-Baille, A., & Chang, N. L. (2020). Techno-economic analysis of hydrogen electrolysis from off-grid stand-alone photovoltaics incorporating uncertainty analysis. *Cell reports Physical science*, 1(10). <https://doi.org/10.1016/j.xcrp.2020.100209>
- [11] Terlouw, T., Bauer, C., McKenna, R., & Mazzotti, M. (2022). Large-scale hydrogen production via water electrolysis: a techno-economic and environmental assessment. *Energy & Environmental Science*, 15(9), 3583–3602. <https://doi.org/10.1039/d2ee01023b>
- [12] Ma, Z., Witteman, L., Wrubel, J. A., & Bender, G. (2021). A comprehensive modeling method for proton exchange membrane electrolyzer development. *International Journal of Hydrogen Energy*, 46(34), 17627–17643. <https://doi.org/10.1016/j.ijhydene.2021.02.170>
- [13] Sakas, G., Ibáñez-Rioja, A., Ruuskanen, V., Kosonen, A., Ahola, J., & Bergmann, O. (2022). Dynamic energy and mass balance model for an industrial alkaline water electrolyzer plant process. *International Journal of Hydrogen Energy*, 47(7), 4328–4345. <https://doi.org/10.1016/j.ijhydene.2021.11.126>
- [14] Brauns, J., & Turek, T. (2022). Experimental evaluation of dynamic operating concepts for alkaline water electrolyzers powered by renewable energy. *Electrochimica Acta*, 404, 139715. <https://doi.org/10.1016/j.electacta.2021.139715>
- [15] Santos, A. L., Cebola, M.-J., & Santos, D. M. F. (2021). Towards the Hydrogen Economy—A Review of the Parameters That Influence the Efficiency of Alkaline Water Electrolyzers. *Energies*, 14(11), 3193. <https://doi.org/10.3390/en14113193>
- [16] Caparrós Mancera, J. J., Segura Manzano, F., Andújar, J. M., Vivas, F. J., & Calderón, A. J. (2020). An Optimized Balance of Plant for a Medium-Size PEM Electrolyzer: Design, Control and Physical Implementation. *Electronics*, 9(5), 871. <https://doi.org/10.3390/electronics9050871>
- [17] Scheepers, F., Stähler, M., Stähler, A., Rauls, E., Müller, M., Carmo, M., & Lehnert, W. (2021). Temperature optimization for improving polymer electrolyte membrane-water electrolysis system efficiency. *Applied Energy*, 283, 116270. <https://doi.org/10.1016/j.apenergy.2020.116270>
- [18] Liu, Z., Zhu, Z., Gao, J., & Xu, C. (2021). Forecast Methods for Time Series Data: A Survey. *IEEE Access*, 9, 91896–91912. <https://doi.org/10.1109/access.2021.3091162>
- [19] Choi, K., Yi, J., Park, C., & Yoon, S. (2021). Deep Learning for Anomaly Detection in Time-Series Data: Review, Analysis, and Guidelines. *IEEE Access*, 9, 120043–120065. <https://doi.org/10.1109/access.2021.3107975>
- [20] Kim, D.-K., & Kim, K. (2022). A Convolutional Transformer Model for Multivariate Time Series Prediction. *IEEE Access*, 10, 101319–101329. <https://doi.org/10.1109/access.2022.3203416>
- [21] Zhao, Z., Xia, C., Chi, L., Chang, X., Li, W., Yang, T., & Zomaya, A. Y. (2021). Short-Term Load Forecasting Based on the Transformer Model. *Information*, 12(12), 516. <https://doi.org/10.3390/info12120516>
- [22] Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., Young, J.-G., & Petri, G. (2020). Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports*, 874, 1–92. <https://doi.org/10.1016/j.physrep.2020.05.004>
- [23] Zhao, D., He, Q., Yu, J., Jiang, J., Li, X., & Ni, M. (2020). Dynamic behaviour and control strategy of high temperature proton exchange membrane electrolyzer cells (HT-PEMECs) for hydrogen production. *International Journal of Hydrogen Energy*, 45(51), 26613–26622. <https://doi.org/10.1016/j.ijhydene.2020.07.155>
- [24] Ogumerem, G. S., & Pistikopoulos, E. N. (2020). Parametric optimization and control for a smart Proton Exchange Membrane Water Electrolysis (PEMWE) system. *Journal of Process Control*, 91, 37–49. <https://doi.org/10.1016/j.jprocont.2020.05.002>
- [25] Toubeau, J.-F., Bottieau, J., Wang, Y., & Vallee, F. (2022). Interpretable Probabilistic Forecasting of Imbalances in Renewable-Dominated Electricity Systems. *IEEE Transactions on Sustainable Energy*, 13(2), 1267–1277. <https://doi.org/10.1109/tste.2021.3092137>

- [26] Liao, W., Bak-Jensen, B., Pillai, J. R., Wang, Y., & Wang, Y. (2021). A review of graph neural networks and their applications in power systems. *Journal of Modern Power Systems and Clean Energy*, 10(2), 345-360. <https://doi.org/10.35833/mpce.2021.000058>
- [27] Biggio, L., & Kastanis, I. (2020). Prognostics and health management of industrial assets: Current progress and road ahead. *Frontiers in Artificial Intelligence*, 3, 578613. <https://doi.org/10.3389/frai.2020.578613>
- [28] Kong, X., Jiang, X., Zhang, B., Yuan, J., & Ge, Z. (2022). Latent variable models in the era of industrial big data: Extension and beyond. *Annual Reviews in Control*, 54, 167–199. <https://doi.org/10.1016/j.arcontrol.2022.09.005>
- [29] Zhan, J., Wu, C., Ma, X., Yang, C., Miao, Q., & Wang, S. (2022). Abnormal vibration detection of wind turbine based on temporal convolution network and multivariate coefficient of variation. *Mechanical Systems and Signal Processing*, 174, 109082. <https://doi.org/10.1016/j.ymssp.2022.109082>
- [30] Chadha, G. S., Shah, S. R. B., Schwung, A., & Ding, S. X. (2022). Shared temporal attention transformer for remaining useful lifetime estimation. *Ieee Access*, 10, 74244-74258. <https://doi.org/10.1109/ACCESS.2022.3187702>
- [31] Theissler, A., Spinnato, F., Schlegel, U., & Guidotti, R. (2022). Explainable AI for time series classification: a review, taxonomy and research directions. *Ieee Access*, 10, 100700-100724. <https://doi.org/10.1109/access.2022.3207765>
- [32] Afrasiabi, M., Mohammadi, M., Rastegar, M., Stankovic, L., Afrasiabi, S., & Khazaei, M. (2020). Deep-based conditional probability density function forecasting of residential loads. *IEEE Transactions on Smart Grid*, 11(4), 3646-3657. <https://doi.org/10.1109/tsg.2020.2972513>