Semarak Engineering Journal

Journal homepage:
https://semarakilmu.my/index.php/sej/index
ISSN: 3036-0145

A Hybrid Framework Integrating Meta-Analytic Evidence, Physics-Informed Deep Learning, and Monte Carlo Simulation for Risk-Aware Green Hydrogen Design

Mochamad Subchan Mauludin[1,*], Arif Rifan Rudiyanto[1], Singgih Dwi Prasetyo[2], Yuki Trisnoaji[2]

[1] Department of Informatics Engineering, Universitas Wahid Hasyim, Semarang 50236, Indonesia
[2] Power Plant Engineering Technology, State University of Malang, Malang 65145, Indonesia

ARTICLE INFO

Article history:
Received 12 March 2026
Received in revised form 21 April 2026
Accepted 12 July 2026
Available online 5 August 2026

Keywords:
Green hydrogen; Hybrid renewable energy; Physics-informed deep learning; Monte Carlo simulation; Conditional Value-at-Risk; Meta-analysis

ABSTRACT

Green hydrogen supplied by variable solar and wind resources is exposed to coupled technical, economic, environmental, and operational uncertainty that deterministic optimization can conceal. This study develops a four-layer decision framework with explicit data contracts: structured evidence synthesis produces parameter priors, bounds, correlations, and provenance; a physics-informed deep-learning (PIDL) surrogate maps operating states to physically admissible electrolyzer-specific electricity consumption; Latin-hypercube Monte Carlo simulation propagates the joint priors through an hourly rule-based dispatch model; and Conditional Value-at-Risk (CVaR) ranks a predefined set of candidate designs by expected and tail cost. The numerical demonstration uses one synthetic renewable year, literature-derived cost distributions, and synthetic PIDL observations; it therefore verifies internal behavior rather than real-world or site-specific performance. Within this illustrative benchmark, PIDL reduced test RMSE from 0.70 to 0.21 kWh kg-1 and eliminated out-of-domain test predictions. Across 10,000 scenarios, the lowest-cost candidate in the tested set achieved a mean LCOH of USD 4.62 kg-1 and a 95% CVaR of USD 6.49 kg-1, compared with USD 5.08 and USD 7.16 kg-1 for the designated risk-neutral case. Financing costs, electrolyzer capital costs, and wind resource uncertainty dominated the variability. Because these costs remain above typical unabated fossil-hydrogen benchmarks, early projects would require a combination of de-risked finance, long-term offtake, demand creation, production support, and/or carbon-value instruments. The results do not include plant validation, battery aging, or explicit logistics for hydrogen storage, transport, and delivery.

1. Introduction

Green hydrogen is increasingly positioned as an energy carrier for decarbonizing ammonia production, refining, high-temperature industrial processes, heavy transport, and long-duration energy storage. Its climate value, however, depends on how electricity is produced, how efficiently the electrolyzer is operated, and whether infrastructure is sized and dispatched to absorb variable

* *Corresponding author.*
E-mail address: aan.subhan18@unwahas.ac.id

https://doi.org/10.37934/sej.15.1.125148

renewable generation. Staffell et al. [1] emphasized that hydrogen is best understood as part of a coupled energy system rather than as an isolated fuel. At the same time, Glenk and Reichelstein [2] showed that the temporal value of electricity and electrolyzer utilization shapes renewable-to-hydrogen economics. Hybrid photovoltaic (PV)-wind portfolios can improve complementarity and reduce extended low-generation events, but the resulting design problem remains highly nonlinear and uncertain.

Electrolyzer technology adds another layer of complexity. Proton-exchange-membrane (PEM) units offer rapid response and broad operating flexibility. In contrast, alkaline systems generally provide mature, lower-cost conversion at the expense of distinct dynamic and gas-purity constraints. Reviews by Carmo et al. [3], Zeng and Zhang [4], and Ursúa et al. [5] document how temperature, current density, pressure, degradation, auxiliary loads, and partial-load operation affect efficiency and durability. Consequently, a nominal efficiency or a single levelized cost of hydrogen (LCOH) cannot adequately represent plant performance across weather years, market conditions, equipment states, and financing structures.

Four methodological streams have advanced the field, but are often applied as parallel analyses rather than as a traceable chain. Systematic review and meta-analysis can characterize between-study heterogeneity; deep learning can approximate the behavior of renewables and electrolyzers; physics-informed learning can constrain extrapolation; and stochastic risk analysis can propagate uncertainty in costs and resources. Recent work has begun to combine meta-analytical evidence with deep learning for hybrid renewable hydrogen systems. At the same time, broader reviews document the expanding applications of artificial intelligence in PV-wind optimization [6,7]. However, the input-output contracts among evidence synthesis, surrogate prediction, scenario generation, and risk preference remain insufficiently explicit. Consequently, an apparently integrated model can still mix empirical evidence, synthetic assumptions, physical regularization, and decision-maker preferences without showing which layer generated each quantity.

1.1 Research Gap and Contribution

The central gap is not the absence of evidence synthesis, artificial intelligence, stochastic simulation, or risk optimization individually, but the absence of a traceable interface among them. Literature-derived values are often transferred into engineering models as unqualified point estimates; neural surrogates are trained without explicit conservation, monotonicity, or admissibility constraints; Monte Carlo inputs are specified without retaining evidence quality or dependence assumptions; and tail-risk metrics are reported without showing how they alter the design decision. This fragmentation weakens reproducibility and can yield apparently precise decisions that are valid only for the calibration dataset, the selected weather year, or the assumed cost range.

This study therefore proposes a hybrid framework that (i) converts heterogeneous literature and project evidence into credibility-weighted probability distributions, bounds, and correlation structures; (ii) embeds electrochemical and energy-balance knowledge into a PIDL surrogate; (iii) propagates aleatory and epistemic uncertainty through Latin-hypercube Monte Carlo simulation; and (iv) evaluates candidate capacity and dispatch configurations using expected performance and CVaR. The interaction is sequential and non-circular. The evidence layer supplies priors $p(\theta|E)$ and applicability metadata; the PIDL layer supplies a scenario-callable response function and physical-residual diagnostics; the scenario layer combines sampled θ, renewable chronology, dispatch, and the surrogate to produce scenario losses $L(x,\xi)$; and the CVaR layer aggregates those losses to rank only the candidate set being tested. CVaR does not retrain the surrogate, and simulation outputs do not become evidence unless an explicit updating protocol is applied. The numerical values are synthetic and literature-informed; they are presented as an internal consistency demonstration, not as a site-specific feasibility study, real-time controller, or universally optimal design.

Table 1
Research gaps addressed by the proposed hybrid framework

Observed limitation	Decision consequence	Framework response	Verification output
Literature values are transferred as point estimates.	Underestimated parameter uncertainty and hidden study heterogeneity	Random-effects evidence synthesis and credibility-weighted priors	Pooled distributions, I^2, prediction intervals, provenance record
Data-driven surrogate unconstrained by physics.	Implausible extrapolation and weak transfer to unseen operating states	Physics residuals, boundary constraints, and admissibility penalties	RMSE, conservation residual, violation rate, calibration error
Deterministic or mean-value design	Fragile capacity choices under weather, cost, and degradation uncertainty	Latin-hypercube Monte Carlo scenario engine	LCOH distribution, reliability, emissions, Value-at-Risk
Average-risk optimization only	Severe downside exposure is hidden by favorable mean performance	CVaR-constrained multi-objective optimization	Mean-CVaR Pareto frontier and risk budget
Interfaces among analytical layers are implicit.	Evidence, surrogate error, scenario uncertainty, and risk preference can be conflated or double-counted.	Typed, one-way data contracts from evidence priors to PIDL outputs, scenario losses, and CVaR ranking.	Provenance map, interface variables, uncertainty ownership, and layer-specific validation.
Illustrative simulations are presented without an explicit claim boundary.	Synthetic numerical performance may be mistaken for field validation or commercial feasibility.	Declared numerical boundary, excluded logistics, and staged empirical validation requirements.	Synthetic-data label, external-validation plan, scope exclusions, and decision-use restrictions.

2. Literature Review

The source synthesis that motivated this study groups prior research into six closely related streams: evidence synthesis and environmental assessment; hybrid renewable energy and electrolyzer modeling; data-driven and physics-informed prediction; stochastic uncertainty propagation; risk-aware and multi-objective optimization; and deployment issues involving explainability, interoperability, regulation, and digital twins. The streams are individually mature, yet their assumptions, data structures, and validation criteria are rarely integrated into a single auditable decision process. This section, therefore, reviews not only what each stream contributes but also how its unresolved limitations motivate the integrated framework developed in Section 3.

The review is organized as a structured critical synthesis rather than a new bibliometric count. Foundational electrolysis, hydrogen-system, and energy-modeling studies are combined with recent work on stochastic investment assessment, deep learning, physics-informed machine learning, life-cycle harmonization, and dynamic electrolyzer control. The analytical emphasis is placed on five questions: what decision is supported, which physical and economic variables are represented, how uncertainty is characterized, how model validity is established, and whether outputs can be transferred from long-term planning to operational control.

2.1 Literature Taxonomy and System Context

Green hydrogen systems couple weather-dependent generation, power electronics, electrochemical conversion, compression, storage, and end-use demand. Reviews of hydrogen in the global energy system show that its value is strongest where direct electrification is difficult and where hydrogen can provide feedstock, seasonal balancing, or transportable energy [1,8]. Economic studies further demonstrate that the competitiveness of renewable hydrogen depends jointly on the cost of renewable electricity, electrolyzer capital cost, utilization, financing, and the ability to operate flexibly when electricity has low opportunity cost [2,9]. Consequently, optimizing a single component without accounting for its interactions with the rest of the system can yield technically feasible yet economically misleading designs.

Hybrid photovoltaic-wind portfolios are important because resource complementarity can reduce prolonged low-power periods, improve electrolyzer utilization, and lower storage requirements. Nevertheless, complementarity is site-specific and scale-dependent. Hourly chronology, the correlation between wind and solar output, curtailment rules, and the electrolyzer's minimum-load and start-stop behavior all influence the preferred capacity mix. General energy-system modeling guidance therefore recommends temporally explicit models, transparent boundary definitions, and sensitivity analysis rather than reliance on annual-average capacity factors alone [10,11]. Recent hybrid-system studies also show how site-specific resource complementarity, deep-learning prediction, and meta-analytical evidence can be combined, while optimization-focused reviews caution that artificial-intelligence gains depend on data quality, objective definition, and validation strategy [6,7,12].

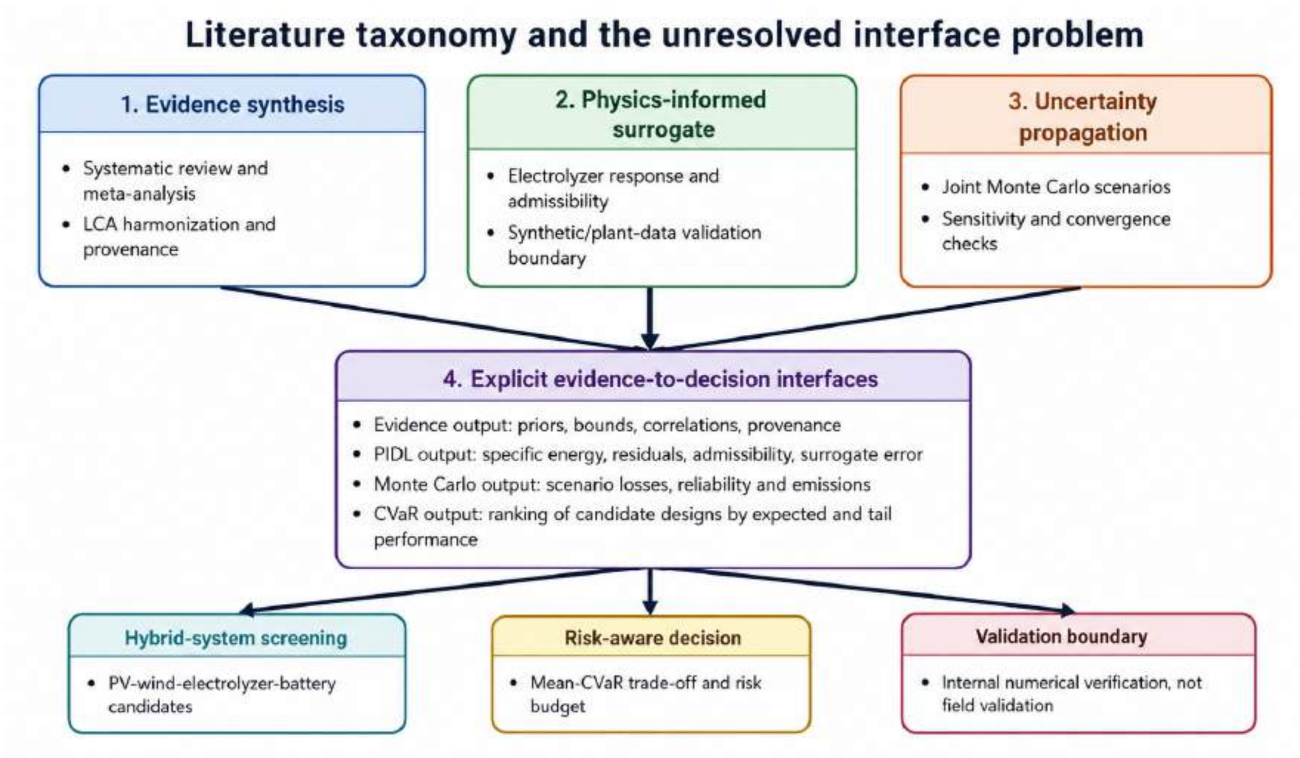

Fig. 1. Literature taxonomy emphasizing explicit interfaces and the numerical validation boundary

Figure 1 summarizes the literature as an evidence-to-decision chain. Systematic review and life-cycle evidence define defensible parameter distributions; physical and data-driven models translate weather and operating states into hydrogen output and equipment stress; stochastic simulation propagates aleatory and epistemic uncertainty; and a risk-aware decision layer compares expected and tail outcomes. The principal research gap is the weak traceability between these layers. For example, a cost range extracted from literature may be used in a Monte Carlo model without retaining study quality, applicable scale, currency year, or correlation with efficiency and lifetime. The revised taxonomy, therefore, emphasizes explicit interface variables and validation boundaries rather than claiming that the current numerical demonstration implements operational deployment.

2.2 Hybrid Renewable Systems and Electrolyzer Technology

Water electrolysis literature distinguishes alkaline, proton-exchange-membrane (PEM), and solid-oxide pathways by maturity, operating temperature, dynamic flexibility, materials, efficiency, and capital intensity. Alkaline electrolyzers are commercially established and can offer lower equipment cost, but conventional configurations generally impose tighter minimum-load and start-up constraints. PEM systems offer compact design, high current density, rapid response, and compatibility with fluctuating renewable power output, although catalyst and membrane costs remain significant [3-5,13]. Technology selection should therefore be treated as a system-level choice rather than a fixed input, because the value of dynamic flexibility depends on resource variability, storage configuration, electricity market participation, and degradation costs.

Electrolyzer models range from constant specific-energy coefficients to semi-empirical polarization curves, electrochemical-thermal models, and multi-stack dynamic models. Constant-efficiency models are computationally convenient for capacity planning but obscure part-load losses, temperature dependence, gas crossover, auxiliary loads, and degradation. Detailed first-principles models improve physical interpretation but can become too slow for thousands of hourly scenarios or nested optimization. Reviews of PEM modeling therefore recommend a hierarchy in which model fidelity is matched to the decision scale, with explicit validation of voltage, Faradaic efficiency, thermal dynamics, and specific electricity consumption [3,14]. Recent grid-integration reviews similarly identify a persistent mismatch between fast system-level dispatch models and the slower thermo-electrical states that constrain real electrolyzers [15].

Storage and balance-of-plant assumptions are equally consequential. Batteries can smooth short-duration fluctuations in renewable output, whereas hydrogen storage and transport decouple production from delivery over longer periods. Compression pressure, drying and purification, storage leakage, inventory cycling, pipeline or trucking distance, and delivery profiles affect energy consumption, reliability, and delivered cost. Supply-chain studies show that the preferred balance among local production, storage, and alternative carriers changes with distance, seasonality, and demand reliability [10]. Delivered-cost analysis further demonstrates that storage and distribution can materially narrow the set of economically attractive end uses [16]. In the present demonstration, however, these hydrogen logistics elements are outside the dynamic optimization boundary: conditioning is represented schematically, and their costs are not resolved by pressure level, storage duration, transport mode, or delivery reliability.

2.3 Meta-Analytic Evidence, Life-Cycle Assessment, and Harmonization

Evidence synthesis is necessary because published values for electrolyzer cost, efficiency, lifetime, renewable capacity factor, water consumption, and life-cycle emissions are heterogeneous.

The heterogeneity is not simply statistical noise: studies differ in technology year, plant scale, currency basis, learning assumptions, system boundaries, electricity source, compression pressure, replacement schedule, and geographical context. Systematic review and random-effects meta-analysis provide a disciplined way to document these differences, quantify residual between-study variance, and produce prediction intervals rather than a single pooled estimate [17-20]. For engineering decisions, the most useful output is therefore a distribution with provenance and applicability metadata, rather than a single point estimate.

Life-cycle assessment (LCA) studies demonstrate the consequences of inconsistent boundaries. Earlier reviews found wide variation in reported environmental performance across hydrogen production methods and emphasized the influence of electricity mix, functional unit, allocation, compression, infrastructure, and water treatment [21,22]. Harmonization studies subsequently showed that recalculating published cases under common assumptions can materially improve comparability and reveal where apparent technology differences are actually methodological artifacts [23]. A recent systematic review of 100 hydrogen-system LCA studies combined qualitative review with meta-regression and showed that a limited set of methodological and technical variables explains a substantial share of reported greenhouse-gas variability [24].

The literature also reveals a separation between environmental assessment and optimization. Standalone LCAs often use detailed cradle-to-grave inventories but provide limited operational resolution, whereas optimization studies frequently reduce environmental performance to a single carbon-intensity coefficient. Reviews of LCA integrated with multi-objective optimization identify missing end-of-life stages, limited impact categories, and insufficient treatment of prospective technology change [25]. The implication for the proposed framework is that environmental parameters must be harmonized before optimization and then propagated as uncertain, scenario-dependent quantities alongside cost and reliability.

2.4 Deep Learning and Physics-Informed Modeling

Deep learning has entered renewable hydrogen research primarily through short-term weather and power forecasting, hydrogen output prediction, anomaly detection, and surrogate modeling. Convolutional, recurrent, and hybrid architectures can capture nonlinear temporal relationships that are difficult to express through fixed regression. Localized wind and solar forecasting has been used to support hydrogen production scheduling, while decomposition and gated recurrent architectures have been applied directly to green hydrogen prediction [26,27]. Recent PV-wind research has also combined deep learning with particle swarm optimization under tropical conditions and reviewed the applications of artificial intelligence in hybrid renewable plants [7,12]. These studies demonstrate predictive and optimization potential, but conventional networks remain sensitive to training-domain coverage, sensor quality, distribution shift, and the physical plausibility of extrapolated outputs.

Physics-informed machine learning addresses this weakness by embedding conservation laws, constitutive relations, known monotonicity, boundary conditions, or differential-equation residuals into the learning objective. The general PINN literature shows that combining sparse observations with governing physics can improve data efficiency and constrain inverse or high-dimensional problems [28-30]. In an electrolyzer application, physics can be introduced at different levels: voltage decomposition, energy conservation, temperature evolution, Faradaic efficiency, pressure bounds, or admissible part-load behavior. The physics term should not be treated as automatically correct; model structural error, poorly scaled residuals, and conflicting loss terms can degrade training.

Validation must therefore report prediction error and physical-residual or constraint-violation metrics.

The broader renewable-hydrogen modeling literature recommends hybrid models in which representative simulation, machine-learning prediction, multi-objective optimization, and decision analysis are selected according to the resolution required by the application [31]. Physics-informed surrogates are particularly attractive for nested Monte Carlo optimization because they can approximate a higher-fidelity electrochemical model at much lower computational cost. Explainability remains important: feature attribution methods can show whether predictions are driven by load, temperature, degradation state, or spurious correlations, thereby supporting operator acceptance and model governance [32].

2.5 Uncertainty Quantification and Risk-Aware Optimization

Techno-economic assessments often report a deterministic levelized cost of hydrogen even though the underlying variables are uncertain and correlated. Capital expenditure, weighted average cost of capital, electricity price, stack life, efficiency, utilization, and replacement cost can jointly change investment outcomes. Recent studies have used Monte Carlo simulation to obtain distributions of LCOH and financial risk, demonstrating that deterministic estimates can conceal asymmetric downside exposure [33,34]. The distinction between aleatory variability, such as weather and market fluctuations, and epistemic uncertainty, such as limited knowledge of future capital cost or degradation, is important because the two categories call for different mitigation strategies.

Monte Carlo analysis is only as credible as its input distributions and dependence structure. Latin-hypercube sampling improves marginal coverage relative to simple random sampling for a fixed scenario budget [35], but it does not remove the need to justify bounds, tails, and correlations. Sensitivity analysis should accompany simulation to identify which uncertainties dominate output variance and which data-collection or contracting actions have the highest value [36,37]. Scenario convergence, out-of-sample testing, and stress cases for prolonged renewable drought, high financing cost, and early stack replacement are necessary to avoid false precision.

Risk-neutral optimization minimizes expected cost, whereas project developers and lenders are also concerned with low-probability, high-loss outcomes. Conditional Value-at-Risk (CVaR) provides a coherent tail-risk metric and can be embedded in linear or nonlinear optimization [38]. In hydrogen systems, the risk measure can be applied to LCOH, unserved hydrogen, cash flow, or penalties for carbon-intensity non-compliance. A mean-CVaR formulation makes the decision-maker's risk preference explicit. It allows comparison of designs that have similar expected cost but different exposure to adverse weather, prices, or equipment performance.

2.6 Multi-Objective, Multi-Timescale, and Deployment Integration

Renewable-hydrogen planning is inherently multi-objective. Cost, greenhouse-gas emissions, renewable curtailment, water use, reliability, electrolyzer utilization, degradation, and market revenue cannot generally be optimized simultaneously without trade-offs. Pareto methods are preferable to a single weighted score when stakeholder weights are uncertain because they preserve the shape of the decision frontier. Recent reviews show that many studies remain bi-objective and use limited environmental detail, reinforcing the need to combine richer LCA indicators with computationally efficient surrogates [25,31].

A second integration problem concerns the time scale. Capacity decisions span decades, whereas dispatch decisions occur hourly or faster. Advanced studies may use rolling horizons, scenario trees, or model predictive control to connect planning and operations [15]. The present numerical demonstration does not implement those methods. It uses a deterministic, myopic hourly rule set without look-ahead. Accordingly, rolling-horizon or MPC comparison is treated as outside the current scope rather than as an untested implication of the reported results.

Model transfer beyond an illustrative numerical study requires disciplined data and governance. At a minimum, timestamps, units, asset identifiers, sensor-quality flags, uncertainty metadata, and model versions must be recorded to enable reconstruction of evidence inputs, surrogate outputs, sampled scenarios, and decision criteria. These are validation and reproducibility requirements; the present study does not claim a deployed digital twin, supervisory controller, or cybersecurity architecture.

2.7 Comparative Synthesis and Research Gap

Table 2 compares representative contributions across the literature streams. The comparison indicates that prior studies usually achieve depth within one or two layers but leave at least one critical interface unresolved. Electrolyzer reviews provide physical realism but do not propagate literature heterogeneity into investment risk. LCA and meta-regression improve comparability of evidence but rarely operate at hourly dispatch resolution. Deep-learning studies improve forecasting but generally lack tail-risk objectives, while stochastic cost studies often rely on simplified or fixed-efficiency electrolysis models.

Table 2
Comparative synthesis of representative literature and the residual integration gap

Study	Primary focus and method	Treatment of physics or uncertainty	Principal contribution	Residual gap relative to this study
Staffell et al., [1]	System-level hydrogen review	Broad scenario evidence; no unified stochastic model	Clarifies cross-sector roles of hydrogen	Does not connect evidence synthesis to operational optimization
Glenk and Reichelstein [2]	Renewable-to-hydrogen economics	Economic sensitivities and operational value	Links the renewable electricity opportunity cost to hydrogen economics	Limited electrochemical dynamics and environmental uncertainty
Buttler and Spliethoff [13]	Electrolysis for storage and sector coupling	Technology and integration constraints reviewed	Establishes system role and operating characteristics	No evidence-to-prior, AI, or tail-risk architecture
Bhandari et al., [21]	Review of hydrogen LCA	Highlights methodological heterogeneity	Shows the importance of the electricity mix and boundaries	Environmental evidence is not coupled to dispatch or finance risk
Benalcazar and Komorowska [33]	Techno-economic uncertainty assessment	Probabilistic cost and export-country scenarios	Demonstrates uncertainty-sensitive LCOH	Uses no physics-informed surrogate or meta-analytic evidence layer
Andrade et al., [34]	Monte Carlo investment risk	Stochastic LCOH and risk metrics	Shows divergence between deterministic and risk-based results	Physical and environmental models remain comparatively aggregated
Nikulins et al., [26]	Deep-learning renewable forecast	Data-driven forecast uncertainty	Improves short-term wind and solar information for H2 control	No physical constraints or investment-risk optimization
Emeksiz and Tan [27]	Hybrid deep-learning H2 prediction	Nonlinear time-series prediction	Demonstrates advanced recurrent forecasting	Limited lifecycle, explainability, and tail-risk treatment

Study	Primary focus and method	Treatment of physics or uncertainty	Principal contribution	Residual gap relative to this study
Karniadakis et al., [29]	Physics-informed machine learning	Physical laws embedded in learning	Provides a general theory for data-physics integration	Not a hydrogen-specific evidence and decision framework
Puig-Samper et al., [24]	LCA systematic review and meta-regression	Explains cross-study GHG variability	Transforms heterogeneous LCA evidence into quantitative drivers	Does not integrate dynamic operation, PIDL, or CVaR
Mukelabai et al., [31]	Renewable-H2 modeling and ML review	Method hierarchy and decision framework	Synthesizes the best modeling and optimization practices	No explicit evidence-to-distribution and tail-risk pipeline
Larrahondo Chávez et al., [25]	LCA and multi-objective optimization review	Environmental-economic trade-offs	Identifies limited LCA depth and end-of-life treatment	Limited dynamic electrolyzer and stochastic operational integration
Yuan et al., [15]	Dynamic modeling and grid control review	Thermo-electrical states and control constraints	Identifies overestimated flexibility in system-level models	No meta-analytic prior, environmental synthesis, or financial tail risk
Trisnoaji et al. [6]	Meta-analysis and deep learning for hybrid PV-wind hydrogen	Evidence synthesis and predictive modeling	Shows an emerging combined evidence-AI pathway	Does not explicitly separate scenario risk and CVaR decision preference
Mauludin et al. [7]	Literature review of AI for hybrid PV-wind optimization	AI methods and optimization applications	Maps AI opportunities and methodological trends	No evidence-to-prior and tail-risk interface
Mauludin et al. [12]	Deep learning and particle-swarm optimization for tropical PV-wind prediction	Data-driven prediction and metaheuristic optimization	Demonstrates coupled prediction-optimization workflow	No hydrogen logistics, physical regularization, or CVaR layer

The unresolved gap is therefore architectural. A defensible design process must preserve the chain from extracted evidence to probability distributions, constrain machine-learning predictions with electrochemical and energy-balance knowledge, propagate joint uncertainty through chronologically explicit simulation, and evaluate capacity and dispatch candidates using both expected performance and tail risk. The framework in Section 3 supplies this chain through explicit inputs, outputs, uncertainty ownership, and validation tests. It remains modular so that site data or higher-fidelity logistics models can replace illustrative components without changing the meaning of upstream evidence or downstream risk preference.

3. Methodology

3.1 Integrated Framework Architecture

Figure 2 organizes the proposed workflow into four linked analytical layers. Layer 1, structured evidence, receives peer-reviewed studies, life-cycle inventories, vendor quotations, operating records, standards, and explicitly labeled expert assumptions. After harmonization, it exports a machine-readable parameter object containing the distribution family, location, and scale parameters, admissible bounds, dependence assumptions, evidence date, applicable technology and scale, and credibility score. These outputs are the only literature-derived inputs passed to subsequent layers; optimization results do not overwrite them.

Layer 2, the PIDL surrogate, receives operating-state variables such as electrolyzer load fraction, stack temperature, and health state, together with physical constraints that are conceptually distinct from the statistical priors. It exports predicted specific electricity consumption, a physical-residual score, an admissibility flag, and, where estimated, predictive uncertainty. In the numerical demonstration, the surrogate is trained and tested using synthetic observations generated around a compact algebraic reference. Consequently, its error reduction demonstrates the behavior of regularization under controlled conditions and cannot be interpreted as validation against a commercial PEM electrolyzer.

Layer 3, Monte Carlo simulation, samples a joint vector θ from the evidence priors, applies one renewable chronology and the hourly dispatch rule for each candidate design x, calls the PIDL surrogate at the resulting operating states, and computes scenario-specific hydrogen output, curtailment, LCOH, reliability, and environmental indicators. The resulting sample {L(x,ξs)} is a distribution of conditional outcomes, not a new evidence base. Surrogate error, when quantified, should be treated as an additional sampled term rather than silently absorbed into broad engineering factors.

Layer 4, CVaR evaluation, consumes only the scenario-loss distribution. At confidence level α, it reports the Value-at-Risk threshold and the average loss beyond that threshold. In a full optimization, these quantities can appear either in the objective or in the constraints. In the present numerical demonstration, they are used to screen a predefined set of representative configurations; no claim of global optimality is made. The only feedback shown in Figure 2 is a governance loop: sensitivity and evidence-confidence results may prioritize future measurements, but priors are updated only when new documented evidence or plant data are supplied.

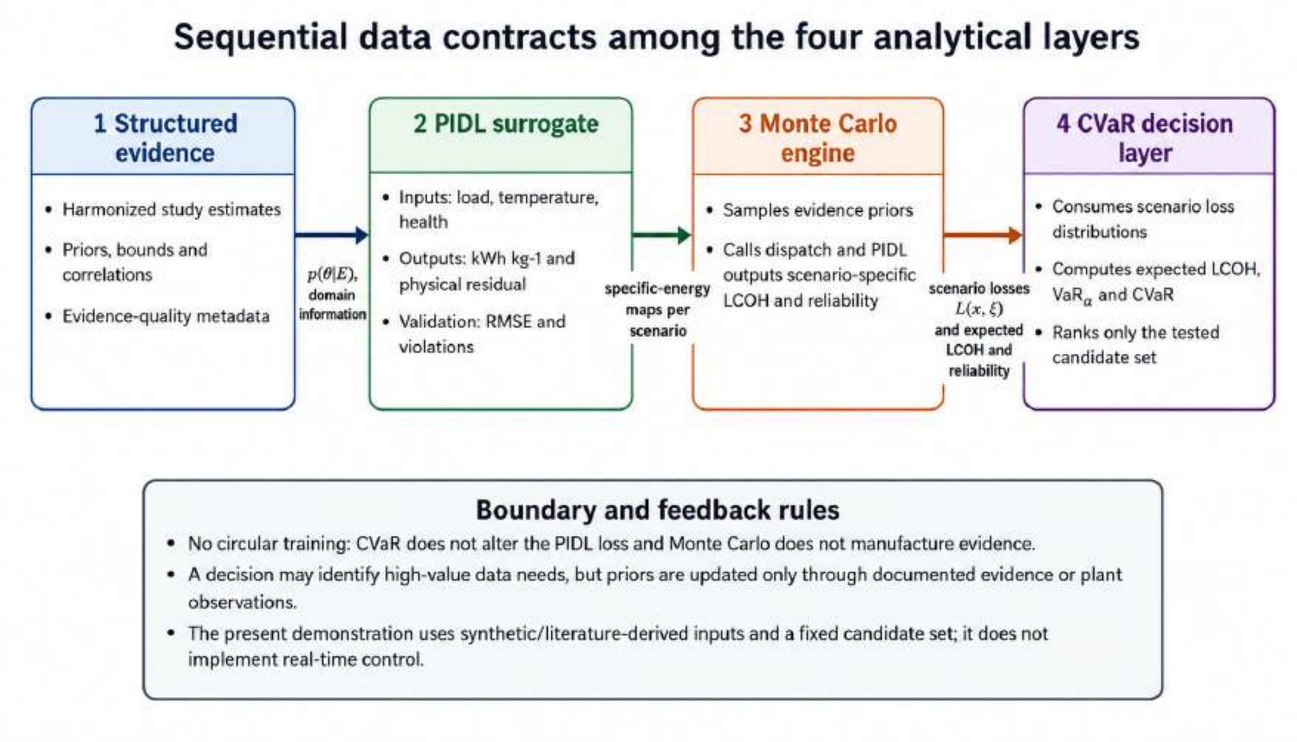

Fig. 2. Sequential data contracts among structured evidence, PIDL, Monte Carlo simulation, and CVaR evaluation

3.2 Meta-Analytic Evidence Layer

The evidence-layer protocol follows the reporting logic of PRISMA 2020 [17] but is adapted to engineering parameters. Search strings combine hydrogen, electrolysis, hybrid renewable systems, techno-economic analysis, life-cycle assessment, forecasting, degradation, and uncertainty. Each study is indexed by technology, geography, year, system boundary, currency basis, functional unit, temporal resolution, data origin, and validation status. Monetary values are normalized to a common price year; efficiency is converted to a consistent lower-heating-value or higher-heating-value basis; and environmental indicators are mapped to kg CO2-equivalent per kg H2. The present article specifies this protocol and uses literature-informed ranges in Table 3, but it does not report an executed, registered systematic review with a released screening log. A future empirical application must release the complete query, inclusion decisions, extracted dataset, and harmonization code.

For a scalar parameter, the study-level estimate y_i with within-study variance v_i is combined using a random-effects model. The between-study variance τ^2 captures residual heterogeneity that remains after harmonization. The pooled estimate and its standard error are computed using inverse-variance weights. At the same time, the I^2 statistic reports the proportion of total variability attributable to heterogeneity rather than sampling error, following DerSimonian and Laird [19] and Higgins et al. [20]. Because engineering studies often report ranges rather than standard errors, the framework permits distribution fitting from minimum-mode-maximum values, confidence intervals, or digitized sensitivity outputs. Still, it records a lower credibility score for indirectly reconstructed data.

$$w_i = \frac{1}{v_i + \tau^2}, \quad \hat{\mu} = \frac{\sum w_i\, y_i}{\sum w_i}, \quad \mathrm{Var}(\hat{\mu}) = \frac{1}{\sum w_i} \tag{1}$$

$$I^2 = \max\left[0, \frac{Q-(k-1)}{Q}\right] \times 100\% \tag{2}$$

The synthesized output is not reduced to a single mean. It becomes a prior distribution p(θ|E) accompanied by source provenance, credibility weight, and applicable domain. Correlations are retained only when physically or economically justified, for example, between electrolyzer capital cost and efficiency, wind resource and curtailment, or interest rate and weighted average cost of capital. Bayesian updating could revise a prior with plant measurements D to produce p(θ|E,D), but that update is not performed in the numerical demonstration. This separation allows decision-makers to distinguish among published evidence, local data, physical constraints, surrogate uncertainty, and scenario assumptions.

Table 3
Evidence-informed uncertainty specification used in the numerical demonstration

Parameter	Illustrative distribution	Unit	Interpretation and update pathway
PV capital cost	Triangular (500, 650, 850)	USD kW-1	Broad literature- and vendor-informed range; update with local EPC quotations
Wind capital cost	Triangular (1050, 1250, 1600)	USD kW-1	Includes balance-of-plant uncertainty; updated by turbine class and site logistics
PEM electrolyzer capital cost	Triangular (600, 800, 1150)	USD kW-1	Consistent with the wide cost dispersion reported in electrolysis cost reviews [9]
Battery capital cost	Triangular (180, 260, 420)	USD kWh-1	Optional short-duration storage; update by chemistry, duration, and warranty
Weighted average cost of capital	Uniform (4, 12)	%	Represents finance and country-risk sensitivity
Plant life	Discrete uniform (20, 30)	year	Used in capital recovery and degradation averaging
Availability	Uniform (94, 99)	%	Joint effect of planned and forced outage

Parameter	Illustrative distribution	Unit	Interpretation and update pathway
Hydrogen efficiency multiplier	Triangular (0.94, 1.00, 1.06)	dimensionless	Aggregates stack efficiency, auxiliary consumption, and calibration uncertainty
PV resource multiplier	Truncated normal (1.00, 0.08)	dimensionless	Interannual and model uncertainty
Wind resource multiplier	Truncated normal (1.00, 0.10)	dimensionless	Interannual and model uncertainty
Fixed operation and maintenance	Uniform (1.8, 3.5)	% of CAPEX y-1	Includes routine plant and balance-of-system cost
Stack replacement fraction	Triangular (0.20, 0.28, 0.40)	fraction of electrolyzer CAPEX	Represents replacement scope and price uncertainty

3.3 Hybrid Renewable Hydrogen System Model

Figure 3 distinguishes the modeled numerical boundary from the broader hydrogen supply chain. The modeled boundary includes PV and wind generation, aggregate power conversion losses, a simplified battery state-of-charge balance, and a PEM electrolyzer represented by the PIDL specific-energy map. At each hour, a deterministic myopic rule first assigns available renewable power to the electrolyzer, charges the battery with any surplus power, and discharges the battery when renewable power is insufficient and within power and state-of-charge limits. The rule has no forecast look-ahead, terminal-value function, market bidding, reserve provision, or MPC optimization. The electrolyzer is shut down if the load falls below the 8% minimum stable load, and the remaining surplus is curtailed.

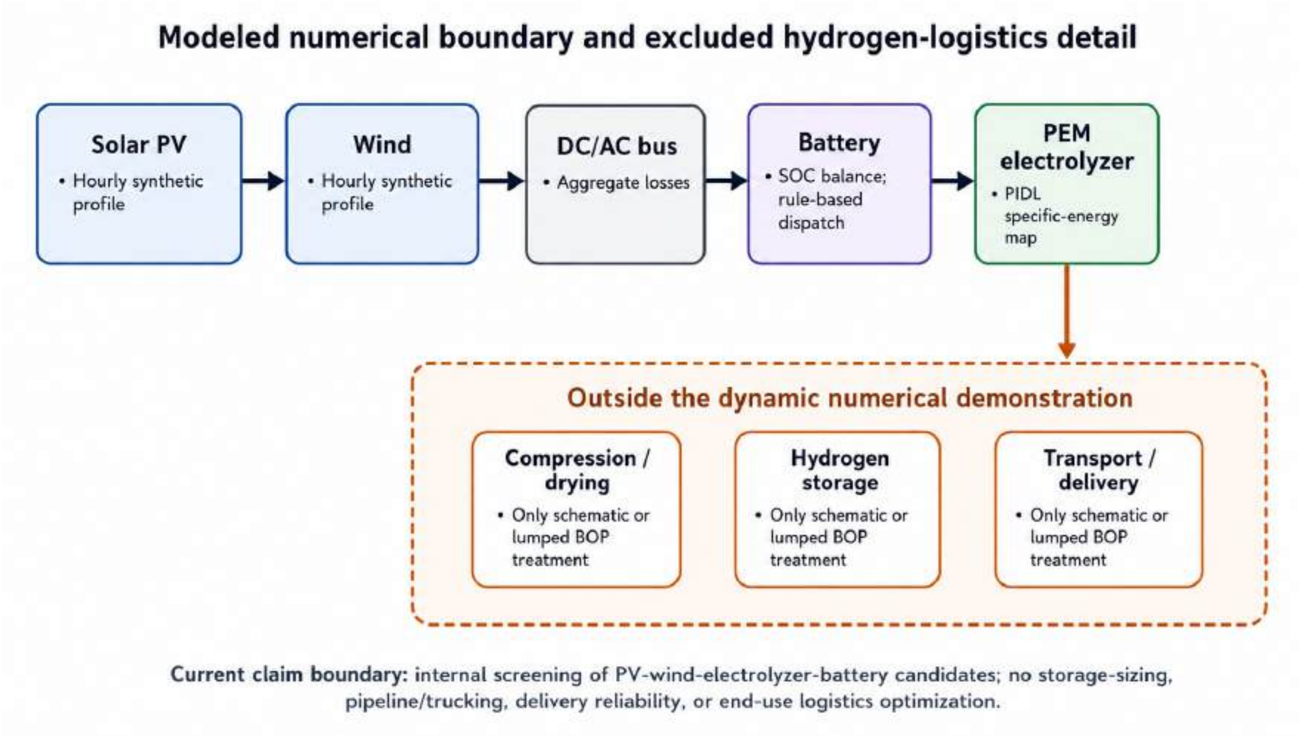

Fig. 3. Modeled numerical boundary and hydrogen-storage/transport elements excluded from dynamic optimization

The battery model is intentionally simplified. It uses aggregate charge/discharge efficiency and state-of-charge limits but omits electrochemical aging, temperature-dependent power limits, cycle-life cost, state-of-health feedback, and replacement optimization. Hydrogen conditioning, storage, transport, and delivery are shown only to define the physical context. Compression stages, purity specifications, storage pressure and leakage, inventory dynamics, pipeline or trucking distance, offtake variability, and delivery reliability are not dynamically modeled. Consequently, the reported LCOH is a production-gate screening metric rather than a delivered-hydrogen cost.

$$P_{\mathrm{RE}}(t) = P_{\mathrm{PV}}(t) + P_{\mathrm{W}}(t), \quad 0 \leq P_{\mathrm{EL}}(t) \leq P_{\mathrm{EL,rated}} \tag{3}$$

$$\mathrm{SOC}(t+1) = \mathrm{SOC}(t) + \eta_{\mathrm{c}} P_{\mathrm{c}}(t)\Delta t - \frac{P_{\mathrm{d}}(t)\Delta t}{\eta_{\mathrm{d}}} \tag{4}$$

$$m_{\mathrm{H_2}}(t) = \frac{P_{\mathrm{EL}}(t)\Delta t}{e_{\mathrm{spec}}[\lambda(t), T(t), h(t)]} \tag{5}$$

The cost model annualizes PV, wind, electrolyzer, battery, and aggregate balance-of-plant capital using the capital recovery factor. Annual operating cost and periodic stack replacement are added to the numerator, while degradation-adjusted hydrogen production forms the denominator. The environmental module can accept full life-cycle inventories, but the demonstration uses a compact embodied-equipment model solely to illustrate multi-objective processing. Neither the cost nor the environmental impact includes a resolved hydrogen storage and transport chain. A project application must align system boundaries with the intended hydrogen certification scheme and include pressure-specific conditioning, storage duration, transport mode, and delivered-demand reliability [16,21,22].

$$\mathrm{LCOH} = \frac{\mathrm{CRF} \times \mathrm{CAPEX} + \mathrm{OPEX} + C_{\mathrm{rep}} + C_{\mathrm{grid}}}{M_{\mathrm{H_2,annual}}} \tag{6}$$

3.4 Physics-Informed Deep-Learning Surrogate

The PIDL surrogate estimates electrolyzer-specific electricity consumption as a function of load fraction and stack temperature. A feed-forward network is trained on 55 sparse, noisy synthetic observations generated around an algebraic reference, with the loss including both a data term and a physics term, evaluated at 500 collocation points. The reference imposes the expected nonlinear efficiency penalty at low load, a shallow optimum near the rated region, and a bounded temperature response. It is not a calibrated electrochemical model and does not resolve polarization, thermal inertia, gas crossover, pressure dynamics, or degradation. The general structure follows physics-informed learning principles [28-30], but the numerical performance reported here is a controlled benchmark rather than plant validation.

$$\mathcal{L} = \lambda_{\mathrm{data}} \mathrm{MSE}(y, \hat{y}) + \lambda_{\mathrm{phys}} \mathrm{MSE}\left(R_{\mathrm{phys}}\right) + \lambda_{\mathrm{bound}} P_{\mathrm{bound}} + \lambda_{\mathrm{cons}} P_{\mathrm{cons}} \tag{7}$$

A valid surrogate must be evaluated beyond ordinary prediction error. The proposed protocol reports RMSE, MAE, physical residual, prediction-interval coverage, and the percentage of outputs that violate known admissible bounds. Temporal validation of real plant data should use blocked or rolling-origin splits to prevent leakage, while domain-shift tests should deliberately evaluate uncommon low-load and high-temperature states. Explainability can be added after validation to test whether load, temperature, and health variables affect predictions in directions consistent with engineering knowledge [32]. Because the current test set is synthetic and generated from the same reference family as the training data, it cannot assess robustness to sensor bias, equipment-to-equipment variation, or out-of-family degradation.

3.5 Monte Carlo Simulation and CVaR Optimization

Monte Carlo analysis samples N joint scenarios from the evidence-informed distributions. Latin-hypercube sampling stratifies each marginal distribution and improves space-filling relative to naive random sampling for a fixed N [35]. Correlated samples may be generated through a Gaussian copula or rank-correlation transformation. For each candidate x and scenario ξs, the sampled parameters are passed through the renewable chronology, myopic dispatch model, PIDL surrogate, cost model, and environmental module. The output is a scenario record containing LCOH, hydrogen production, curtailment, reliability, emissions, and diagnostic flags. This explicit record prevents the evidence distribution, surrogate residual, and realized scenario outcome from being treated as interchangeable forms of uncertainty.

CVaR measures the expected loss conditional on outcomes exceeding a selected Value-at-Risk threshold. Unlike a single percentile, CVaR remains sensitive to the magnitude of severe outcomes and can be incorporated into linear or nonlinear optimization [38]. For loss L(x,ξ), design x, uncertainty ξ, and confidence level α, the sample approximation introduces an auxiliary threshold ζ. In a full optimization, expected LCOH and CVaR can be combined in an objective or risk-budget constraint. In this article, CVaR is applied after simulation to compare a predefined set of candidates. It does not modify the PIDL training loss, generate new parameter priors, or prove that the selected case is globally optimal.

$$\mathrm{CVaR}_{\alpha}(L) = \min_{\zeta}\left\{\zeta + \frac{1}{(1-\alpha)N}\sum_{s}\max[L(x,\xi_s)-\zeta,0]\right\} \tag{8}$$

$$\min_{x}\left\{\mathbb{E}(\mathrm{LCOH}), \mathrm{CVaR}_{0.95}(\mathrm{LCOH}), \mathrm{GWP}, U_{\mathrm{H}_2}\right\} \tag{9}$$

Global sensitivity analysis is performed after scenario convergence. Rank correlations provide an interpretable screening metric, while variance-based indices can be added for nonlinear interactions [36,37]. Monte Carlo convergence is assessed by repeating the analysis with independent seeds and monitoring the stability of the means, tail quantiles, CVaR, and candidate ranking. A field application must also propagate calibrated surrogate uncertainty and test structurally adverse cases, such as correlated renewable drought, early stack replacement, restricted battery availability, and storage or delivery bottlenecks. Not all of those stressors are represented in the current demonstration.

3.6 Numerical Demonstration and Validation Protocol

The demonstration uses one synthetic hourly year with a mean PV capacity factor of approximately 0.23 and a wind capacity factor of approximately 0.38. The combined renewable portfolio is fixed at 200 MW, while PV-wind allocation, electrolyzer capacity, and battery energy vary across three representative configurations. The Monte Carlo experiment contains 10,000 Latin-hypercube scenarios and the uncertain variables listed in Table 3. Renewable chronology, PIDL training observations, and the comparison set are synthetic or literature-derived. The case, therefore, tests only numerical consistency, sensitivity, and reporting logic; it does not validate a site, a real electrolyzer, an operating controller, or a delivered hydrogen supply chain.

Table 4
Numerical case configuration and validation design

Element	Specification	Validation or reporting requirement
Temporal model	8,760 hourly steps	Report renewable capacity factors, missing-data treatment, and chronology
Renewable capacity	200 MW total PV plus wind	Evaluate alternative PV-wind shares and weather years
Electrolyzer	110-140 MW PEM	Minimum stable load 8%; temperature and load-dependent efficiency
Battery	0-40 MWh; aggregate SOC and efficiency	Rule-based, no-look-ahead dispatch; battery aging, temperature, market services, and replacement optimization excluded
PIDL data	55 synthetic noisy training points; 1,000 synthetic test points; 500 collocation points	Internal benchmark only; requires independent commercial-stack or pilot-plant validation
Uncertainty engine	10,000 Latin-hypercube scenarios	Repeat with independent seeds; verify tail-metric convergence
Risk metric	95% Value-at-Risk and CVaR	Report both the threshold and the conditional tail average
Outputs	LCOH, annual hydrogen, carbon intensity, curtailment, sensitivity	Separate illustrative results from site-specific evidence
Hydrogen logistics boundary	Conditioning shown schematically; storage and transport not dynamically modeled	Do not interpret production-gate LCOH as delivered cost; add compression, storage, transport, and offtake constraints for project use

4. Results and Discussion

4.1 Evidence-to-Prior Integration

The specified evidence layer is designed to produce a machine-readable parameter registry rather than an unstructured list of literature values. Each record should contain the harmonized unit, central estimate, uncertainty distribution, applicable technology and scale, correlation assumptions, evidence date, and confidence score. In this article, Table 3 illustrates the registry structure using broad literature- and vendor-informed distributions; it is not the output of a completed meta-analysis. Accordingly, the demonstration tests how such priors would be consumed, not whether the listed distributions are definitive for a particular market or site.

The practical value of the architecture is that evidence confidence can be separated from decision sensitivity. In an applied study, a parameter supported by multiple consistent measurements would receive a narrower evidence distribution, whereas an emerging-technology parameter would retain a wider prediction interval. Sensitivity and confidence can then be combined to prioritize data acquisition. This value-of-information logic remains prospective here because no registered evidence dataset or site-measurement campaign is evaluated.

4.2 Physics-Informed Model Performance

Figure 4 compares the data-only and physics-informed networks over electrolyzer load at 60 °C within the synthetic benchmark. With the same sparse, noisy generated training set, the data-only model achieved an RMSE of 0.70 kWh kg-1 and an MAE of 0.53 kWh kg-1. Adding the physics residual reduced RMSE to 0.21 kWh kg-1 and MAE to 0.15 kWh kg-1. The data-only model generated 3.7%

predictions outside the imposed 47.5-65 kWh kg-1 admissible range, while the PIDL model generated 0.0%. These values demonstrate internal regularization behavior under the assumed reference function; they are not measurement errors from a real PEM stack.

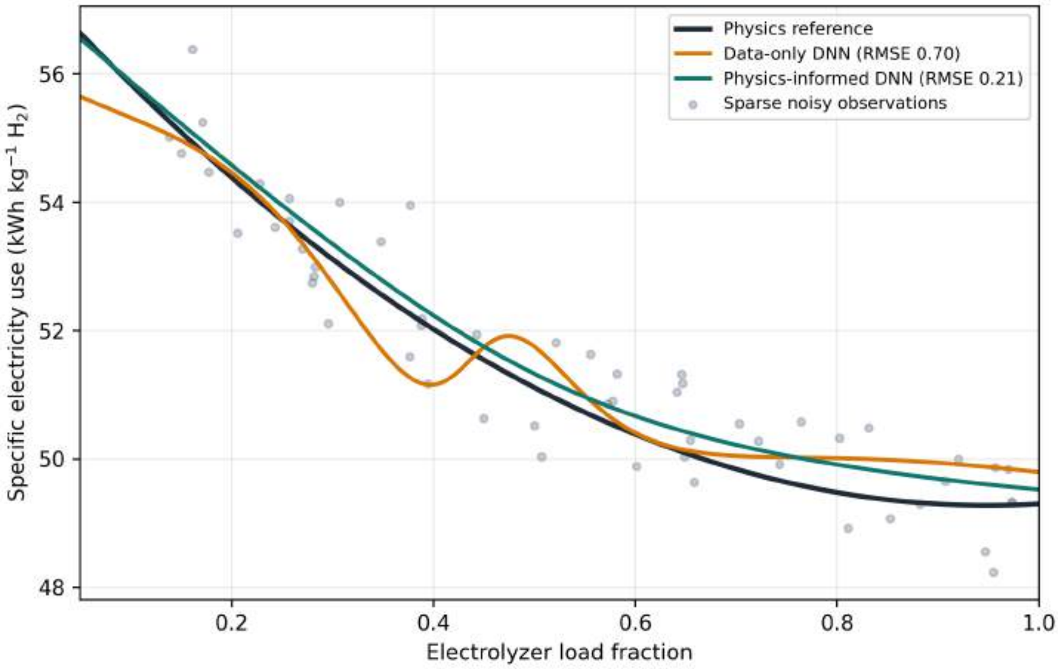

Fig. 4. Data-only and physics-informed predictions of electrolyzer-specific electricity consumption

Table 5
Electrolyzer surrogate validation results

Model	RMSE (kWh kg-1)	MAE (kWh kg-1)	Admissibility violations (%)	Interpretation
Data-only deep neural network	0.703	0.530	3.7	Fits observations but extrapolates poorly at low load
Physics-informed deep neural network	0.206	0.153	0.0	Lower error and physically admissible across the test domain

Within the controlled synthetic benchmark, the result demonstrates why physical regularization can be valuable when data are sparse. It should not be interpreted as proof that a physics-informed model will outperform a data-only model on plant data. Incorrect equations, poorly scaled loss terms, biased sensors, unmodeled degradation, or conflicting constraints can degrade performance. External validation must therefore include independent operating records, equipment-level holdout tests, ablation studies, residual inspection, calibration analysis, and comparison with physically credible baseline models.

4.3 Risk-Aware Design Results

The Monte Carlo LCOH distributions in Figure 5 differ in both center and tail. The designated risk-neutral configuration has a mean LCOH of USD 5.08 kg-1, a 95th percentile of USD 6.72 kg-1, and a 95% CVaR of USD 7.16 kg-1—the balanced hybrid yields USD 4.75, 6.27, and 6.68 kg-1, respectively.

The candidate labeled CVaR-aware has the lowest mean and tail costs among the three tested configurations, with a mean LCOH of USD 4.62 kg-1, USD 6.11 kg-1 at the 95th percentile, and a CVaR of USD 6.49 kg-1. Relative to the designated risk-neutral case, the tail cost falls by approximately 9.4%, and the mean hydrogen output rises from 8.32 to 9.22 kt y-1. The comparison does not establish a global optimum or out-of-sample superiority.

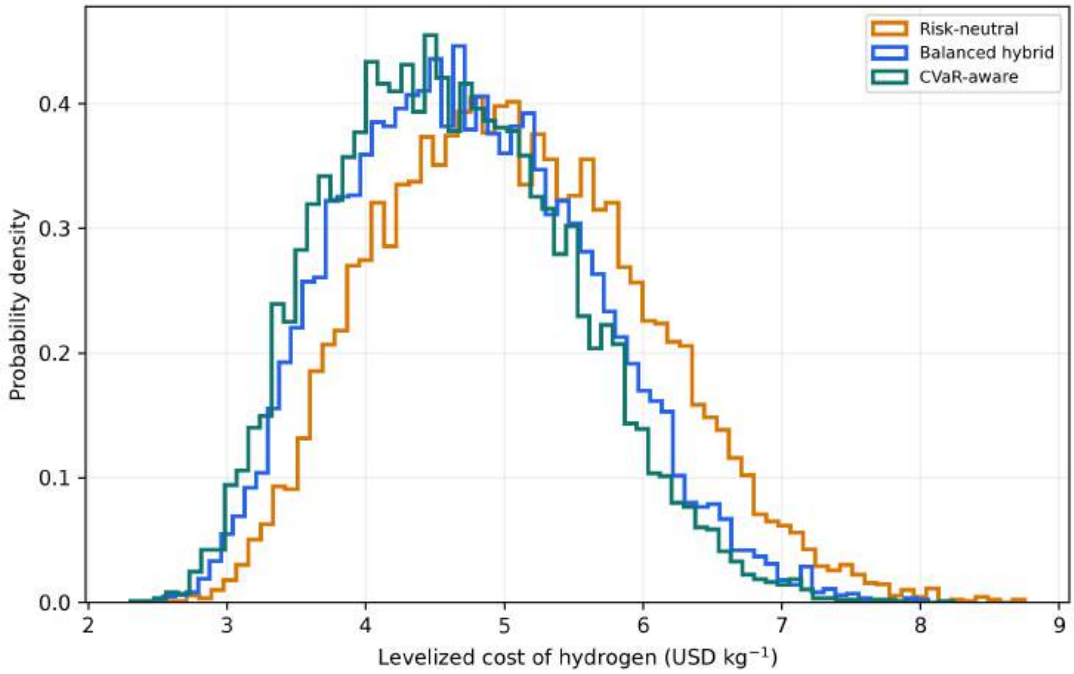

Fig. 5. Monte Carlo distributions of levelized hydrogen cost for three design strategies

Table 6
Monte Carlo performance of the representative design strategies

Strategy	PV	Wind	EL	Battery	Mean LCOH	P95	CVaR95	H2	Curtail.
Risk-neutral	120	80	140	0	5.08	6.72	7.16	8.32	0.48
Balanced hybrid	100	100	120	20	4.75	6.27	6.68	8.77	1.33
CVaR-aware	80	120	110	40	4.62	6.11	6.49	9.22	2.19

Within the assumed synthetic year and simplified dispatch rule, the difference is associated with complementary wind output, higher electrolyzer utilization, and limited battery energy that reduces operation near the minimum-load region. The mechanism should be treated as a hypothesis for field testing because it excludes battery degradation, forecast error, market participation, hydrogen inventory, and delivery obligations. A strategy label, such as risk-neutral or CVaR-aware, denotes the role assigned in the illustrative comparison; it is not evidence that either configuration is the mathematical optimum for a real project.

The carbon-intensity estimates remain close, at approximately 1.04-1.05 kg CO2-equivalent kg-1 H2, because renewable-equipment contributions dominate the compact environmental demonstration and no carbon-intensive grid import is modeled. This narrow spread should not be generalized. Real-life cycle results can vary with manufacturing location, replacement rate, renewable-matching rules, water treatment, compression, storage, transport, and allocation of curtailed electricity [21-25]. Because storage and distribution are excluded, the values are production-boundary indicators rather than delivered-hydrogen footprints.

Semarak Engineering Journal
Volume 15, Issue 1 (2026) 125-148

4.4 Sensitivity and Pareto Trade-Offs

Figure 6 shows rank correlations between uncertain inputs and LCOH for the balanced hybrid case. Weighted average cost of capital is dominant (ρ = 0.80), followed by electrolyzer capital cost (ρ = 0.28), wind-resource multiplier (ρ = -0.26), PV capital cost (ρ = 0.24), PV resource (ρ = -0.21), and fixed operation and maintenance (ρ = 0.20). The signs are physically and economically consistent under the assumed model. The dominance of financing is especially relevant to policy design: concessional debt, guarantees, credible long-term offtake, and revenue stabilization may reduce LCOH risk more directly than small technical-efficiency gains. However, rank correlations from the synthetic experiment do not quantify causal effects or serve as a substitute for project finance due diligence.

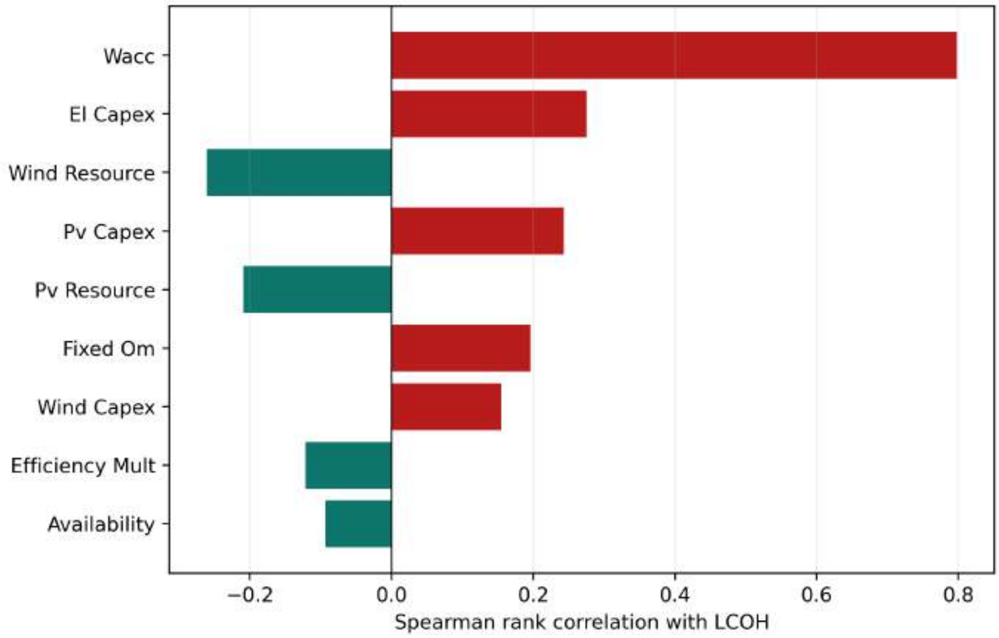

Fig. 6. Spearman rank sensitivity of LCOH for the balanced hybrid strategy

Table 7
Dominant uncertainty drivers and recommended risk-reduction actions

Ranked parameter	Spearman ρ	Recommended project action
Weighted average cost of capital	+0.798	Evaluate concessional finance, guarantees, and bankability of offtake
Electrolyzer capital cost	+0.275	Use vendor quotations and learning scenarios
Wind resource	-0.261	Extend on-site measurement and multi-year reanalysis
PV capital cost	+0.243	Update EPC and supply-chain scenarios
PV resource	-0.209	Validate satellite data with ground measurements
Fixed O&M	+0.196	Benchmark service contracts and local labor/logistics
Wind capital cost	+0.155	Resolve turbine, foundation, and grid-connection scope
Hydrogen efficiency	-0.122	Obtain stack performance maps and auxiliary-load data
Availability	-0.093	Use maintenance and spare parts, reliability models

Figure 7 maps the tested candidates by expected LCOH and CVaR. The non-dominated frontier makes risk preference explicit by showing the exchange between average cost and severe-outcome exposure. The plot is more informative than a single selected point. Still, it remains conditional on the candidate-generation rules, one synthetic weather year, the simplified battery dispatch, and the specified uncertainty distributions. A lender may prefer lower tail risk, while an equity investor may tolerate greater variability; neither preference resolves the omitted risks of storage, transport, delivery, or market.

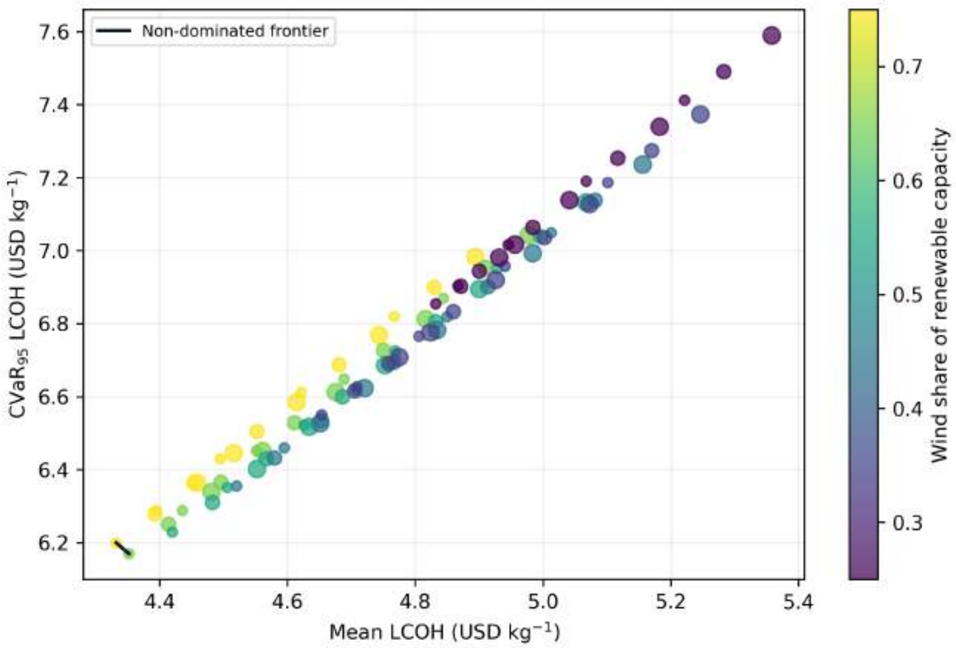

Fig. 7. Mean-CVaR trade-off across candidate PV-wind-electrolyzer-battery designs

4.5 Economic Competitiveness and Policy Implications

The demonstrated production-gate LCOH of USD 4.62 kg-1 at the mean and USD 6.49 kg-1 in the 95% tail remains above commonly reported unabated fossil-hydrogen costs. This means that the numerical improvement from risk-aware configuration selection should not be equated with commercial competitiveness. The International Energy Agency identifies the cost gap between low-emissions and unabated fossil hydrogen as a central barrier and recommends targeted support, demand aggregation, and public procurement during scale-up [39]. Peer-reviewed assessments likewise conclude that substantial policy support may be required and that announced green-hydrogen ambitions exceed projects backed by effective implementation instruments [40,41].

A transparent break-even carbon-price calculation illustrates the magnitude of the gap. Let pCO2* = (LCOHgreen - Cfossil)/(Ifossil - Igreen), where C is production cost, and I is life-cycle emissions intensity. Using a deliberately broad fossil-cost benchmark of USD 1.5-2.5 kg-1, an unabated fossil-hydrogen intensity of 10-12 kg CO2-equivalent kg-1 H2, and the illustrative green intensity of approximately 1.05 kg CO2-equivalent kg-1 H2, the mean LCOH implies roughly USD 190-350 tCO2-equivalent-1, while the 95% CVaR implies roughly USD 360-560 tCO2-equivalent-1. These are scenario calculations, not forecasts: they exclude delivered logistics costs, market-specific taxes, co-product treatment, and certification rules. Their purpose is to show that carbon pricing alone may be insufficient in many markets, especially for tail-risk financing cases.

Table 8, Illustrative policy-bridging requirement for the reported LCOH range. A commercially relevant support package should therefore be layered rather than reduced to a single carbon price number. Production premiums or carbon contracts-for-difference can stabilize the clean-production premium; concessional finance and guarantees can address the dominant WACC sensitivity; long-term offtake, quotas, green public procurement, or product standards can create demand; and credible certification can preserve the carbon value of the product. Carbon contracts for difference are specifically designed to hedge the policy-related carbon price risk of innovative low-carbon investments [42]. Because storage and distribution can add substantial delivered cost, policy appraisal must ultimately be repeated with end-use-specific logistics [16].

Table 8
Policy-Bridging Requirement For The Reported Lcoh Range

Illustrative case	Green LCOH (USD kg-1)	Implied break-even carbon value (USD tCO2e-1)	Interpretation
Mean of the lowest-cost tested candidate	4.62	approximately 190-350	Likely requires combined production support, finance de-risking, and bankable offtake.
95% CVaR of the same candidate	6.49	approximately 360-560	Tail risk materially increases the required support or the certainty of contracted revenue.

4.6 Limitations, Reproducibility, and Focused Research Agenda

The study has six material limitations. First, the parameter distributions are synthetic or literature-informed and were not produced by an executed, registered meta-analysis with released extraction data. Second, the PIDL training and test observations are synthetic, and no real electrolyzer, pilot plant, or commercial operating dataset is used for external validation. Third, the battery is represented by aggregate state-of-charge and efficiency equations under a deterministic myopic rule; aging, temperature, forecast uncertainty, market services, and optimized replacement are omitted. Fourth, hydrogen compression detail, storage inventory, leakage, transport mode and distance, delivery reliability, and offtake logistics are outside the dynamic model, so LCOH should not be interpreted as delivered cost. Fifth, a single synthetic weather year cannot capture multi-decadal climate variability, spatial correlation, or prolonged compound renewable drought. Sixth, three representative configurations and a generated Pareto set do not prove global optimality. These limitations materially restrict site, bankability, and policy conclusions.

Reproducibility should be evaluated at three levels. Evidence reproducibility requires the complete search record, inclusion decisions, extraction sheet, harmonization code, and source-to-prior mapping. Computational reproducibility requires fixed software versions, random seeds, scenario inputs, a synthetic data generator, surrogate hyperparameters, and unit tests for energy and financial balances. Decision reproducibility requires disclosure of candidate-generation rules, objective weights, CVaR confidence level, feasibility tolerances, and stakeholder constraints. Reporting only the final LCOH is insufficient because the same value can arise from different evidence, physical, or risk assumptions.

Future research is organized into two sequential priorities. Priority 1 is empirical validation: preregister and conduct the evidence synthesis, release the harmonized dataset, and validate the PIDL and dispatch outputs using multi-site operational measurements, equipment-level holdouts, and calibrated uncertainty estimates. Priority 2 begins only after that validation boundary is met: extend the model to include battery degradation and explicit constraints on hydrogen compression, storage, transport, and offtake, then recompute delivered cost, reliability, and policy requirements.

This focused sequence replaces unsupported comparison with rolling-horizon MPC and avoids presenting unrelated deployment topics as if they were equally immediate research tasks.

5. Conclusions

This study developed a hybrid framework that connects structured evidence, physics-informed deep learning, Monte Carlo uncertainty propagation, and CVaR evaluation through explicit one-way data contracts. Evidence provides priors and provenance; PIDL provides a constrained response function and physical diagnostics; Monte Carlo converts joint uncertainty into scenario outcomes; and CVaR summarizes severe-cost exposure for a declared candidate set. This architecture clarifies which conclusions are evidence-derived, physically imposed, statistically simulated, or preference-dependent.

Within a synthetic numerical benchmark, physical regularization reduced electrolyzer specific-energy RMSE from 0.70 to 0.21 kWh kg-1 and eliminated violations of the imposed test-domain bounds. Across 10,000 scenarios, the lowest-cost candidate tested achieved a mean LCOH of USD 4.62 kg-1 and a 95% CVaR of USD 6.49 kg-1, versus USD 5.08 kg-1 and USD 7.16 kg-1 for the designated risk-neutral case. Financing cost was the dominant uncertainty driver. These findings verify internal behavior and illustrate how tail-risk awareness can change candidate ranking; they do not constitute plant validation, global optimization, or a bankable project estimate.

The economic result also remains above typical unabated fossil-hydrogen production costs. Under transparent illustrative assumptions, bridging the mean result corresponds to a carbon value of approximately USD 190-350 tCO2-equivalent-1, increasing to approximately USD 360-560 tCO2-equivalent-1 for the 95% tail. This indicates that early deployment is more plausibly supported by a portfolio of carbon-value recognition, production premiums or contracts for difference, concessional finance or guarantees, long-term offtake, demand standards, and credible certification than by technical optimization alone. The immediate research agenda is therefore limited to two steps: first, execute and release the evidence synthesis and validate the surrogate and dispatch model on multi-site operating data; second, add battery degradation and explicit hydrogen storage, transport, and offtake logistics before drawing delivered-cost or policy conclusions.

Acknowledgement
The authors would like to thank Universitas Wahid Hasyim, Semarang, for its institutional support of this research.

Conflict of Interest Statement
The authors declare that there is no conflict of interest regarding the publication of this paper. The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions Statement
The authors' contributions to this research were pivotal in shaping the study's outcomes. Mochamad Subchan Mauludin conceptualized the research framework, led the analysis, and took primary responsibility for writing the manuscript, providing critical insights into integrating meta-analytic evidence with the hybrid decision framework. Arif Rifan Rudiyanto contributed significantly by developing the Monte Carlo simulation methodology, assisting with data analysis, and helping to draft sections focused on the modeling approaches. Singgih Dwi Prasetyo brought his expertise in physics-informed deep learning to the research, playing a key role in developing surrogate modeling

techniques and aiding in interpreting the results. Lastly, Yuki Trisnoaji was instrumental in conducting the numerical simulations, contributing to data collection and analysis, and reviewing the manuscript to enhance its clarity and coherence. Collectively, all authors collaborated closely, reading and approving the final manuscript, ensuring a comprehensive representation of their work.

Data Availability Statement

For access to the available data, please feel free to contact the corresponding author. Your inquiries will be addressed promptly and respectfully. Thank you.

Ethics Statement

This study did not involve human participants, personal data, animals, or biological materials; therefore, institutional ethical approval and informed consent were not required.

References

[1] Staffell, Iain, Daniel Scamman, Anthony Velazquez Abad, Paul Balcombe, Paul E. Dodds, Paul Ekins, Nilay Shah, and Kate R. Ward. 2019. "The Role of Hydrogen and Fuel Cells in the Global Energy System." *Energy & Environmental Science* 12 (2): 463-491. https://doi.org/10.1039/C8EE01157E.

[2] Glenk, Gunther, and Stefan Reichelstein. 2019. "Economics of Converting Renewable Power to Hydrogen." *Nature Energy* 4: 216-222. https://doi.org/10.1038/s41560-019-0326-1.

[3] Carmo, Marcelo, David L. Fritz, Jürgen Mergel, and Detlef Stolten. 2013. "A Comprehensive Review on PEM Water Electrolysis." *International Journal of Hydrogen Energy* 38 (12): 4901-4934. https://doi.org/10.1016/j.ijhydene.2013.01.151.

[4] Zeng, Kai, and Dongke Zhang. 2010. "Recent Progress in Alkaline Water Electrolysis for Hydrogen Production and Applications." *Progress in Energy and Combustion Science* 36 (3): 307-326. https://doi.org/10.1016/j.pecs.2009.11.002.

[5] Ursúa, Alfredo, Luis M. Gandía, and Pablo Sanchis. 2012. "Hydrogen Production from Water Electrolysis: Current Status and Future Trends." *Proceedings of the IEEE* 100 (2): 410-426. https://doi.org/10.1109/JPROC.2011.2156750.

[6] Trisnoaji, Yuki, Singgih Dwi Prasetyo, Zainal Arifin, and Mochamad Subchan Mauludin. 2026. "Enhancing Sustainability of On-Grid Hybrid PV-Wind Systems for Green Hydrogen Production from Unconventional Resources: A Meta-Analytical and Deep Learning Approach." *Fuel* 422: 139157. https://doi.org/10.1016/j.fuel.2026.139157.

[7] Mauludin, Mochamad Subchan, Moh. Khairudin, Rustam Asnawi, Wan Azani Mustafa, and Toha Siti Fauziah. 2025. "The Advancement of Artificial Intelligence's Application in Hybrid Solar and Wind Power Plant Optimization: A Study of the Literature." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 50 (2): 279-293. https://doi.org/10.37934/araset.50.2.279293.

[8] Parra, David, Luca Valverde, F. Javier Pino, and Martin K. Patel. 2019. "A Review on the Role, Cost and Value of Hydrogen Energy Systems for Deep Decarbonisation." *Renewable and Sustainable Energy Reviews* 101: 279-294. https://doi.org/10.1016/j.rser.2018.11.010.

[9] Saba, Sayed M., Martin Müller, Martin Robinius, and Detlef Stolten. 2018. "The Investment Costs of Electrolysis - A Comparison of Cost Studies from the Past 30 Years." *International Journal of Hydrogen Energy* 43 (3): 1209-1223. https://doi.org/10.1016/j.ijhydene.2017.11.115.

[10] Reuß, Markus, Thomas Grube, Martin Robinius, Peter Preuster, Peter Wasserscheid, and Detlef Stolten. 2017. "Seasonal Storage and Alternative Carriers: A Flexible Hydrogen Supply Chain Model." *Applied Energy* 200: 290-302. https://doi.org/10.1016/j.apenergy.2017.05.050.

[11] Pfenninger, Stefan, Adam Hawkes, and James Keirstead. 2014. "Energy Systems Modeling for Twenty-First Century Energy Challenges." *Renewable and Sustainable Energy Reviews* 33: 74-86. https://doi.org/10.1016/j.rser.2014.02.003.

[12] Mauludin, Mochamad Subchan, Moh. Khairudin, Rustam Asnawi, Singgih Dwi Prasetyo, Bimastyaji Surya R., Siti Sendari, and Wan Azani Mustafa. 2026. "Optimizing Energy Production Predictions in Hybrid PV-Wind Systems: Integrating Deep Learning Techniques and Particle Swarm Optimization for Improved Performance in Tropical Climates." *e-Prime - Nexus of Electrical, Electronic, and Intelligent Engineering* 17: 201188. https://doi.org/10.1016/j.eprime.2026.201188.

[13] Buttler, Alexander, and Hartmut Spliethoff. 2018. "Current Status of Water Electrolysis for Energy Storage, Grid Balancing and Sector Coupling via Power-to-Gas and Power-to-Liquids: A Review." *Renewable and Sustainable Energy Reviews* 82: 2440-2454. https://doi.org/10.1016/j.rser.2017.09.003.

[14] Falcão, D. S., and A. M. F. R. Pinto. 2020. "A Review on PEM Electrolyzer Modelling: Guidelines for Beginners." *Journal of Cleaner Production* 261: 121184. https://doi.org/10.1016/j.jclepro.2020.121184.

[15] Yuan, Tiejiang, Jie Tan, and Yue Teng. 2026. "A Review of Dynamic Modeling and Control of Grid-Connected Hydrogen Production Units Using Water Electrolysis." *Renewable and Sustainable Energy Reviews* 226, Part B: 116296. https://doi.org/10.1016/j.rser.2025.116296.

[16] Shafiee, Roxana T., and Daniel P. Schrag. 2024. "Carbon Abatement Costs of Green Hydrogen across End-Use Sectors." *Joule* 8 (12): 3281-3289. https://doi.org/10.1016/j.joule.2024.09.003.

[17] Page, Matthew J., Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, et al. 2021. "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews." *BMJ* 372: n71. https://doi.org/10.1136/bmj.n71.

[18] Borenstein, Michael, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. 2009. *Introduction to Meta-Analysis*. Chichester: Wiley. https://doi.org/10.1002/9780470743386.

[19] DerSimonian, Rebecca, and Nan Laird. 1986. "Meta-Analysis in Clinical Trials." *Controlled Clinical Trials* 7 (3): 177-188. https://doi.org/10.1016/0197-2456(86)90046-2.

[20] Higgins, Julian P. T., Simon G. Thompson, Jonathan J. Deeks, and Douglas G. Altman. 2003. "Measuring Inconsistency in Meta-Analyses." *BMJ* 327: 557-560. https://doi.org/10.1136/bmj.327.7414.557.

[21] Bhandari, Ramchandra, Clemens A. Trudewind, and Petra Zapp. 2014. "Life Cycle Assessment of Hydrogen Production Methods - A Review." *Journal of Cleaner Production* 85: 151-163. https://doi.org/10.1016/j.jclepro.2013.07.048.

[22] Mehmeti, Andi, Alessandro Angelis-Dimakis, George Arampatzis, Stephen McPhail, and Sergio Ulgiati. 2018. "Life Cycle Assessment and Water Footprint of Hydrogen Production Methods: From Conventional to Emerging Technologies." *Environments* 5 (2): 24. https://doi.org/10.3390/environments5020024.

[23] Valente, Antonio, Diego Iribarren, and Javier Dufour. 2018. "Harmonising the Cumulative Energy Demand of Renewable Hydrogen for Robust Comparative Life-Cycle Studies." *Journal of Cleaner Production* 175: 384-393. https://doi.org/10.1016/j.jclepro.2017.12.069.

[24] Puig-Samper, Gonzalo, Eleonora Bargiacchi, Diego Iribarren, and Javier Dufour. 2024. "Life-Cycle Assessment of Hydrogen Systems: A Systematic Review and Meta-Regression Analysis." *Journal of Cleaner Production* 470: 143330. https://doi.org/10.1016/j.jclepro.2024.143330.

[25] Larrahondo Chávez, Diego, Catherine Azzaro-Pantel, Florent Montignac, and Alain Ruby. 2025. "Integrating Life Cycle Assessment in Multi-Objective Optimization of Green Hydrogen Systems: A Review of Literature and Methodological Challenges." *Renewable and Sustainable Energy Reviews* 217: 115689. https://doi.org/10.1016/j.rser.2025.115689.

[26] Nikulins, Arturs, Kaspars Sudars, Edgars Edelmers, Ivars Namatevs, Kaspars Ozols, Vitalijs Komasilovs, Aleksejs Zacepins, Armands Kviesis, and Andreas Reinhardt. 2024. "Deep Learning for Wind and Solar Energy Forecasting in Hydrogen Production." *Energies* 17 (5): 1053. https://doi.org/10.3390/en17051053.

[27] Emeksiz, Cem, and Mustafa Tan. 2024. "An Innovative Adaptive Hybrid Prediction Model Based on Deep Learning Methods (CEEMDAN-GRU) for Green Hydrogen Production: A Case Study of Türkiye." *International Journal of Hydrogen Energy* 52: 1266-1279. https://doi.org/10.1016/j.ijhydene.2023.11.026.

[28] Raissi, Maziar, Paris Perdikaris, and George E. Karniadakis. 2019. "Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations." *Journal of Computational Physics* 378: 686-707. https://doi.org/10.1016/j.jcp.2018.10.045.

[29] Karniadakis, George E., Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. 2021. "Physics-Informed Machine Learning." *Nature Reviews Physics* 3: 422-440. https://doi.org/10.1038/s42254-021-00314-5.

[30] Cuomo, Salvatore, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli. 2022. "Scientific Machine Learning through Physics-Informed Neural Networks: Where We Are and What's Next." *Journal of Scientific Computing* 92: 88. https://doi.org/10.1007/s10915-022-01939-z.

[31] Mukelabai, Mulako D., Edward R. Barbour, and Richard E. Blanchard. 2024. "Modeling and Optimization of Renewable Hydrogen Systems: A Systematic Methodological Review and Machine Learning Integration." *Energy and AI* 18: 100455. https://doi.org/10.1016/j.egyai.2024.100455.

[32] Lundberg, Scott M., Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. 2020. "From Local Explanations to Global Understanding with Explainable AI for Trees." *Nature Machine Intelligence* 2: 56-67. https://doi.org/10.1038/s42256-019-0138-9.

[33] Benalcazar, Pablo, and Aleksandra Komorowska. 2024. "Techno-Economic Analysis and Uncertainty Assessment of Green Hydrogen Production in Future Exporting Countries." *Renewable and Sustainable Energy Reviews* 199: 114512. https://doi.org/10.1016/j.rser.2024.114512.

[34] Andrade, Jorge Vleberton Bessa de, Vinicius Braga Ferreira da Costa, Benedito Donizeti Bonatto, Giancarlo Áquila, Edson de Oliveira Pamplona, and Ramchandra Bhandari. 2024. "Perspective under Uncertainty and Risk in Green Hydrogen Investments: A Stochastic Approach Using Monte Carlo Simulation." *International Journal of Hydrogen Energy* 49: 385-404. https://doi.org/10.1016/j.ijhydene.2023.08.253.

[35] McKay, Michael D., Richard J. Beckman, and William J. Conover. 1979. "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code." *Technometrics* 21 (2): 239-245. https://doi.org/10.2307/1268522.

[36] Saltelli, Andrea, Paola Annoni, Ivano Azzini, Francesca Campolongo, Marco Ratto, and Stefano Tarantola. 2010. "Variance Based Sensitivity Analysis of Model Output: Design and Estimator for the Total Sensitivity Index." *Computer Physics Communications* 181 (2): 259-270. https://doi.org/10.1016/j.cpc.2009.09.018.

[37] Homma, Toshimitsu, and Andrea Saltelli. 1996. "Importance Measures in Global Sensitivity Analysis of Nonlinear Models." *Reliability Engineering & System Safety* 52 (1): 1-17. https://doi.org/10.1016/0951-8320(96)00002-6.

[38] Rockafellar, R. Tyrrell, and Stanislav Uryasev. 2000. "Optimization of Conditional Value-at-Risk." *Journal of Risk* 2 (3): 21-41. https://doi.org/10.21314/JOR.2000.038.

[39] International Energy Agency. 2024. *Global Hydrogen Review 2024*. Paris: IEA. https://www.iea.org/reports/global-hydrogen-review-2024.

[40] Ueckerdt, Falko, Christian Bauer, Alois Dirnaichner, Jordan Everall, Romain Sacchi, and Gunnar Luderer. 2024. "On the Cost Competitiveness of Blue and Green Hydrogen." *Joule* 8 (1): 104-128. https://doi.org/10.1016/j.joule.2023.12.004.

[41] Odenweller, Adrian, and Falko Ueckerdt. 2025. "The Green Hydrogen Ambition and Implementation Gap." *Nature Energy* 10: 110-123. https://doi.org/10.1038/s41560-024-01684-7.

[42] Richstein, Jörn C., and Karsten Neuhoff. 2022. "Carbon Contracts-for-Difference: How to De-Risk Innovative Investments for a Low-Carbon Industry?" *iScience* 25 (8): 104700. https://doi.org/10.1016/j.isci.2022.104700.