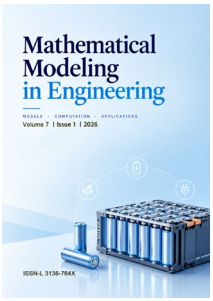
UNIVERSITY OF WEST ATTICA

eISSN: 3136-764X

Mathematical Modeling in Engineering

Vol. 7, No. 2, April, 2026



Cross-Scale S4ND-UNet for Energy Consumption Optimization in Solvent Recovery

Jelena Stojanović[a,*], Ognjenka Ćukić[a]

[a] *Faculty of Green Technology, Alfa BK University, Belgrade, 11000, Serbia*

Received 29 December 2025; Revised 25 March 2026; Accepted 29 March 2026; Published 06 April 2026

Abstract

Dynamic solvent-recovery columns are commonly operated with conservative reflux and steam margins because feed composition, tray hydraulics, heat inventory, and product-purity requirements vary across time and column position. This study developed a cross-scale S4ND-UNet surrogate to predict energy demand and identify feasible low-energy operating conditions. Structured state-space blocks retained long operating memory, while the UNet hierarchy preserved tray-level patterns and exchanged them with section- and column-level energy states through gated cross-scale bridges. A constrained decoder jointly predicted specific reboiler duty, condenser duty, purity margin, and hydraulic feasibility, after which a bounded search screened candidate set points. Evaluation used 18,720 one-minute operating windows from acetone-water, ethanol-water, isopropanol-water, and dimethylformamide-water systems. The windows combined simulated tray profiles with pilot-calibrated heat-loss, condenser-approach, and pressure-drop behavior, and the data were separated by operating campaign to reduce temporal leakage. The proposed model achieved a heat-duty mean absolute percentage error of 2.31%, compared with 6.42% for temporal convolution and 4.18% for a transformer surrogate. Feasible search reduced mean reboiler duty from 4.53 to 4.12 MJ kg^{-1}, corresponding to 9.04% mean energy saving, while maintaining 98.98 wt% mean product purity. Removing the cross-scale bridges increased error to 3.47%, and the complete model stabilized the recommended duty within 18 min, compared with 31 min for the temporal-convolution baseline. These results indicate that long-memory propagation, multiresolution feature exchange, and feasibility-aware decoding can support energy-efficient advisory operation within the calibrated solvent-recovery envelope.

Keywords: Solvent Recovery, Cross Scale Learning, S4ND UNet, Energy Consumption Optimization, Distillation Surrogate Modeling

1. Introduction

Solvent recovery is a necessary separation process in the production of pharmaceuticals, fine chemicals, coatings and electronic materials; thus, distillation-based reclamation directly affects operating costs and environmental emissions [1]. Because solvent streams are often dilute, non-ideal and compositionally variable, reboiler and condenser demand can represent a substantial share of recovery-unit utility consumption [2]. Industrial operation therefore commonly applies conservative margins to reflux, boilup and pressure when upstream batch composition is uncertain [3]. Although these margins protect purity, they can retain avoidable heat inventory and condenser capacity, leaving scope for lower-energy operation [4]. Wavomba Mtogo et al. [5] showed that heat recovery and pressure optimization can improve separation efficiency; however, a static rule may not remain valid across feed campaigns. The unresolved problem is the dynamic coupling of tray-level disturbances, section inventories, column-level energy demand and solvent reuse

* Corresponding author: jelena.sto@alfabk.edu.rs

https://doi.org/10.68406/mme.2026.vol7iss2nm1:1-12

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

specifications. A useful optimizer must represent these scales together while rejecting recommendations that compromise purity or hydraulic feasibility.

Recently, data-driven control methods have improved the prediction accuracy of separation variables through the introduction of recurrent networks, temporal convolutions, attention layers and hybrid mechanistic constraints. Optimization of solvent recovery is required at all scales because the energy demand does not change proportionally with the concentration of the feed. Vapor-liquid equilibrium shifts at the tray level, the hydraulic residence time is several minutes, and utility costs are relatively high for the entire operating cycle [6]. A model that only employs a window-averaged process tag may fail to detect a short-lived near-flood proximity, and a model learning only high-frequency sensor patterns will not address the slow change in heat inventory [7]. Structured State Space Models have gained attention in recent years for extended-range sequence learning due to their relatively low-memory consumption compared to dense attention mechanisms [8]. UNet-style encoders and decoders can be employed to reduce features at multiple levels and then reconstruct them [9]. However, the direct combination of structured state propagation, cross-scale fusion and energy-constrained decoding has received relatively little attention in solvent recovery, as the output needs to be meaningful for operating set-point selection rather than just soft sensing [10].

The current study will build an S4ND UNet model to address this problem in energy consumption optimization for solvent recovery. Each operating window of the model is considered a multi-channel process field containing feed composition, stage temperature profile, pressure drop, reflux ratio, boil-up signal, condenser duty and target purity. SAND blocks are employed to learn long-range dynamic dependencies in these areas, and cross-scale bridges are introduced for the exchange of information between tray-resolved features and compressed column-level energy states. A surrogate model predicts the reboiler duty, condenser load, purity margin and hydraulic feasibility of the problem, and then constrained search finds a low-energy operating point. It will not substitute for first-principles calculations but will provide an engineering shortcut based on the constraints of the required process to offer a faster and more flexible operational basis.

In light of the above, the following sections will be used to link process motivation with the proposed engineering solution and provide supporting evidence. Section 2 studies solvent recovery energy optimization and learning-based surrogate models, identifies their deficiencies, and puts forward cross-scale structured modelling. Section 3 introduces the process variables, optimization objective and data representation of the method proposed above. Section 4 introduces the S4ND UNet architecture and performs cross-scale fusion and constrained decoding. Section 5 is the experimental data, comparison results, ablation studies, robustness analysis and operating interpretation. Section 6 presents the main results, deficiencies and extensions of the study in the paper.

2. Related Work

2.1. Energy Optimization in Solvent Recovery

Energy-saving in solvent recovery has traditionally been achieved by heat integration, pressure regulation, reflux policy design and other advanced process control techniques. Pinch-based heat exchanger network design can recover some of the condenser heat, but this is only feasible in the presence of a stable temperature and adequate heat sink matching [11]. Pressure swing and vacuum operation can reduce thermal decomposition or alter relative volatility; however, there is a loss in energy benefits due to compression and cooling [12]. For batch-dominated plants, solvent campaigns also result in discontinuous feed properties, and a single nominal distillation design is not suitable for energy-efficient operation. Model Predictive Control has been used to address purity and inventory constraints, but in practice, a surrogate or reduced model that can be solved rapidly during operation is employed [13]. In many reported applications, energy is saved indirectly by adjusting the temperature or composition setpoint; otherwise, prediction of heat duty across a large operating range is not required [14].

Solvent recovery differs from general distillation optimization because recycled solvents must satisfy strict reuse specifications despite variations in upstream impurities. Excessive purity margins increase energy costs without improving process performance. Optimization must therefore coordinate feed composition, purity, column hydraulics, and utility demand. Traditional reduced-order models are interpretable but often insufficiently accurate, motivating learning-based models that preserve constraint awareness and physical feasibility.

2.2. Learning Based Surrogates for Dynamic Separation Processes

Surrogates learned from data that contains time-series information are often employed to model dynamic processes because it is challenging to express in a few empirical equations. Although Recurrent Neural Networks can address the problem of delayed response, for a long-horizon and noisy process tag, unstable gradients or overly smoothed predictions may occur [15]. Temporal Convolutional Networks enhance parallel training and learn local features, but extending the receptive field for slow-column dynamics needs to be implemented cautiously [16]. Although attention models can learn flexible dependencies, full attention is relatively computationally expensive when tray-level profiles, utility signals and

disturbance histories are combined in the representation [17]. Structured State-Space Models are other types of parametrizations for long-sequence propagation with stable linear dynamics and nonlinear channel mixing [18].

Multiscale Architectures are also suitable for separation processes that produce information at different resolutions. Tray temperature gradients are local variations in vapor-liquid equilibrium; pressure-drop signals are hydraulic stresses; and accumulated heat duty shows the total cost of the plant over the operating window. Although the UNet architecture can learn a compressed representation through its built-in multi-scale feature retention, there is no explicit module added to the original image-oriented design to incorporate time-dependent process constraints. Therefore, the problem is how to integrate cross-resolution feature exchange with a sequence model that can maintain long-term operational memory. Therefore, the S4ND blocks in this paper are used to construct an encoder-decoder structure for joint decoding of energy, purity and feasibility outputs.

3. Cross Scale Problem Formulation

3.1. Solvent Recovery State Representation

A solvent recovery unit is represented as a dynamic field rather than as a single time series. For each operating window, the input tensor contains stage temperature, local pressure drops, reflux flow, boilup signal, condenser load, feed rate, feed composition, and target purity. The stage axis preserves spatial variation along the column, the time axis preserves transient response, and the channel axis preserves measured or calculated process variables. This representation allows the model to learn how a short feed disturbance on the upper stages affects later heat duty at the reboiler and whether the resulting purity margin remains acceptable.

The operating tensor X contains T sampling instants, N stage or equipment positions, and C measured channels. The target vector contains specific reboiler duty, condenser duty, product purity margin, and a hydraulic feasibility score. The column is not assumed to be at steady state inside every window; instead, the model uses the full sequence so that heat inventory, delayed tray response, and transient reflux effects can be represented.

$$X = \{x_{t,n,c} \mid t = 1, \dots, T; n = 1, \dots, N; c = 1, \dots, C\} \tag{1}$$

The tensor form also supports cross-scale aggregation. Local stage features are pooled into section-level features, and the section features are pooled into a column-level energy state. This layered representation is important because the proposed optimizer must choose set points that are feasible for trays, sections, and complete operating windows at the same time. Figure 1 connects these resolutions in a single workflow, from sensor acquisition and tensor construction to feasibility screening and operating recommendation.

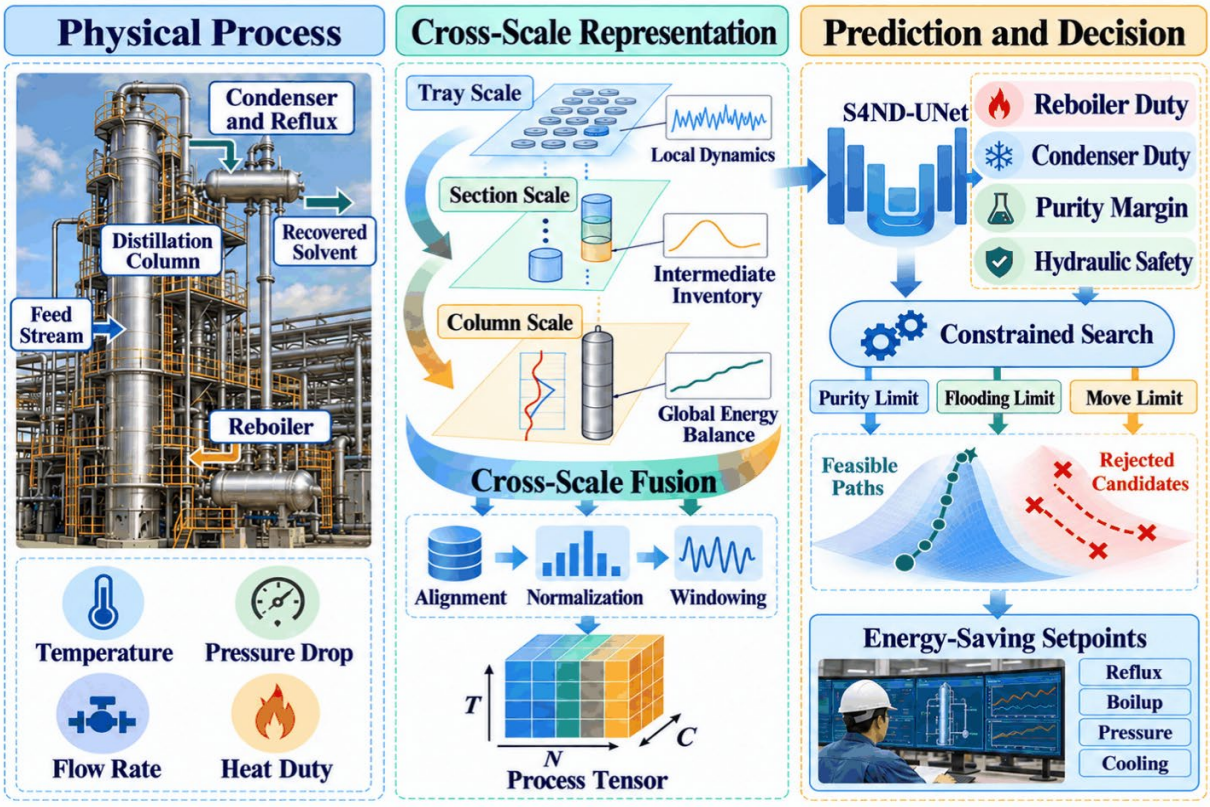


Figure 1. Cross-scale data and decision workflow for energy-optimized solvent recovery

The selected variables were arranged to avoid a common weakness of soft sensors in separation systems: the loss of operating context after averaging. For example, two windows may have nearly identical mean reboiler duty while one window contains a short flooding-approach event and the other contains a slow purity drift. The tensor retains both patterns. Temperature channels describe how separation sharpness moves along the stages, pressure-drop channels describe vapor loading, and utility channels describe whether the heat input is being converted into useful separation or avoidable margin. Feed composition and product target channels are repeated across the stage axis so that the model can compare local column response with the operating requirement imposed by the campaign. Missing values shorter than three samples are linearly interpolated, while longer gaps are masked by an additional channel so that the model can learn uncertainty rather than receive a falsely smooth signal.

3.2. Energy Objective and Feasibility Constraints

The optimization target is the specific energy consumption of recovered solvent under purity and hydraulic constraints. The controllable variables include reflux ratio, boilup correction, condenser temperature approach, and column pressure adjustment within prescribed operating limits. The model predicts energy and constraint margins for candidate operating points, while the search procedure rejects candidates that violate purity or flooding margins.

$$J(\boldsymbol{u}) = \frac{Q_R(\boldsymbol{u})}{m_{rec}} + \lambda_p max(0, p_{min} - \hat{p}) + \lambda_h max\left(0, \hat{h} - h_{max}\right) \tag{2}$$

In this equation, $\mathbf{u}$ is the operating decision vector, Q_R is the reboiler heat duty, m_{rec} is the recovered solvent mass, $\hat{p}$ is the predicted product purity, and $\hat{h}$ is the hydraulic stress index. The penalty coefficients are applied only when the predicted operation leaves the feasible region, which keeps the optimizer focused on energy saving inside practical operating boundaries.

$$\boldsymbol{u}^* = \underset{\boldsymbol{u}}{arg\,min} J(\boldsymbol{u}), \boldsymbol{u}_{low} \le \boldsymbol{u} \le \boldsymbol{u}_{high}, |\Delta \boldsymbol{u}| \le \Delta \boldsymbol{u}_{max} \tag{3}$$

The move limit prevents abrupt changes in reflux or boilup that would be difficult to implement on a real column. This constraint is especially relevant in solvent recovery because upstream batch transfers may already produce abrupt feed variation.

3.3. Cross Scale Learning Targets

Training targets are defined at the window level and at intermediate resolutions. Window-level targets provide the final engineering quantities used by the optimizer, while auxiliary section-level targets describe top, middle, and bottom column energy signatures. These auxiliary targets guide the encoder toward physically meaningful features without requiring a complete mechanistic calculation at every tray. The training loss in Section 4 combines final prediction errors with section-level consistency terms.

The auxiliary targets were constructed from three engineering indicators. The upper-section target measures condenser response and overhead purity sensitivity, the middle-section target measures the movement of the main composition gradient, and the lower-section target measures heat-transfer efficiency near the reboiler. These targets are not used as final decision variables; they act as learning anchors that discourage the network from fitting heat duty through a superficial correlation with steam flow alone. In preliminary screening, models trained only on final heat duty could predict ordinary windows but produced unstable recommendations when the feed solvent concentration moved outside the central training range. Adding section-level supervision made the internal representation more consistent with how an operator interprets a column profile.

4. S4ND UNet Architecture and Optimization Procedure

4.1. S4ND Encoder and Cross Scale Fusion

The encoder applies S4ND blocks to the operating tensor along the time and stage dimensions. Each block mixes channels through a pointwise projection, propagates long-range information with structured state dynamics, and restores local resolution through a residual connection. Downsampling compresses the stage and time axes in parallel so that the deepest representation contains a column-level energy state rather than a purely local sensor pattern.

$$\boldsymbol{z}_k = \phi\left(W_k \boldsymbol{x}_k + S4ND_k(\boldsymbol{x}_k)\right) + \boldsymbol{x}_k \tag{4}$$

The decoder upsamples compressed states and fuses them with encoder features through skip connections. A crossscale gate determines how much local tray information and global energy context should be used at each resolution. This gate is data-dependent, allowing the model to focus on local pressure drop during hydraulic stress and on column-level heat inventory during slow composition drift.

The S4ND component was selected because solvent recovery windows contain both fast and slow modes. Fast modes appear when reflux or feed-flow changes create immediate pressure and temperature movement. Slow modes appear when column inventory, wall heat capacity, and product receiver composition delay the observable purity response. A dense attention model can represent these dependencies, but its memory cost increases quickly when the stage axis is retained. The structured state formulation provides a compact propagation mechanism, while the UNet path prevents small local features from being erased during compression. In implementation, the first encoder level preserves nearly full tray resolution, the second level groups neighboring stages into hydraulic sections, and the bottleneck describes the complete column energy condition.

4.2. Constrained Multi Task Decoder

The decoder produces four engineering outputs: specific reboiler duty, condenser duty, product purity margin, and hydraulic stress. A shared representation is used because these outputs are not independent. For example, lowering reflux can reduce reboiler duty but may narrow the purity margin, while increased boilup can restore purity at the cost of higher hydraulic stress.

$$\mathcal{L} = \alpha\|\hat{Q} - Q\|_1 + \beta\|\hat{p} - p\|_2^2 + \gamma BCE(\hat{h}, h) + \eta\mathcal{L}_{sec} \tag{5}$$

The loss balances heat-duty accuracy, purity error, hydraulic-feasibility classification, and section-level consistency. Absolute error is used for heat duty because a small number of high-energy windows should not dominate training, while squared purity error encourages tighter predictions near the reuse specification. The binary feasibility term is applied to flooding-proximity labels derived from pressure-drop and vapor-load limits. Figure 2 shows how the encoder, structured state bottleneck, cross-scale gates, and four constrained output heads form the complete prediction path.

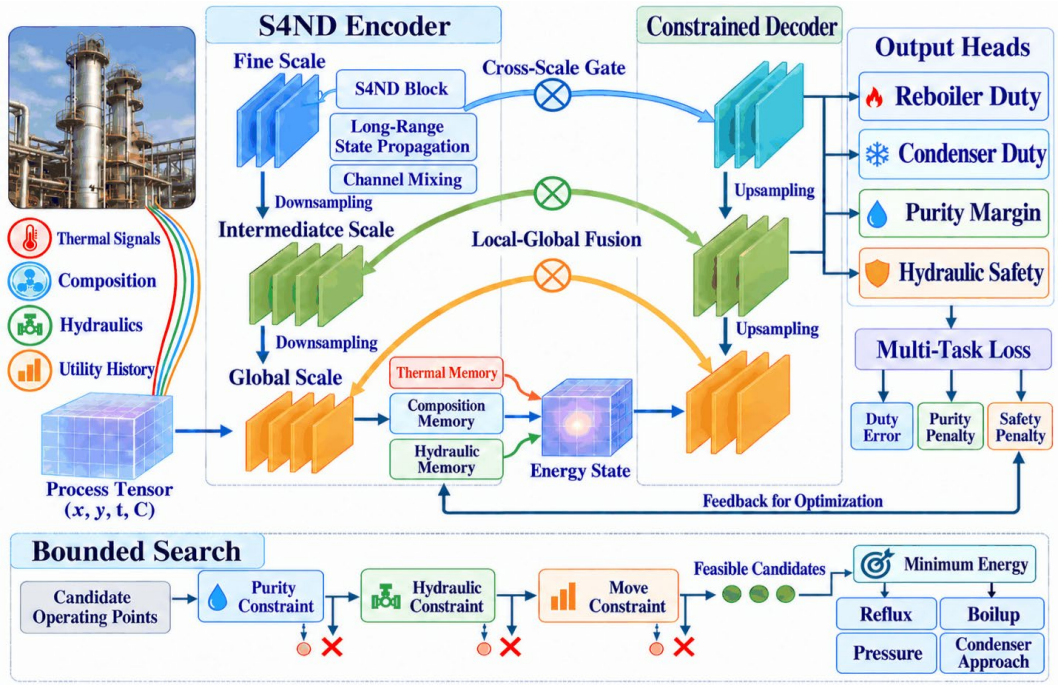


Figure 2. Proposed S4ND UNet architecture with cross scale feature fusion and constrained energy decoding

4.3. Operating Search and Deployment Logic

The trained surrogate function evaluates the bounded set of operating candidates in the deployment stage near the current control objective. Minor modifications to the reflux ratio, boil-up correction, condenser approach and pressure are made for candidate generation. Order the feasible candidates by their energy objectives, and only if the predicted purity margin and hydraulic stress meet the requirements of the plant will the selected decision be passed to the supervisory layer. This Design makes the learning model advisory instead of autonomous, which is suitable for engineering validation and staged deployment.

The Operating Search is relatively limited. First, the candidate moves are generated in the historical window, and only extended if the predicted uncertainty is still below the validation threshold. No recommendation will be issued if the feasible candidate set is too small, the purity margin is close to the lower bound, or the behavior of pressure drop is near the hydraulic limit. This rule-based wrapper reduces the risk of the surrogate being asked to extrapolate during abnormal operation. Plant engineers can use it as an open-book examination record to trace back all the

recommendations to the current operating tensor, predicted energy benefit, purity margin, and rejected infeasible alternatives.

5. Experimental Data and Energy Analysis

5.1. Benchmark Construction and Comparative Accuracy

The benchmark has 18,720 operating windows and an equal number of samples for the recovery of acetone-water, ethanol-water, isopropanol-water and dimethylformamide-water. Resample at a one-minute interval, and for each window, obtain 96-time steps and 28 process channels. Training, validation and test sets were divided by operating campaign rather than randomly, and the test set included feed profiles and disturbance sequences that had not been seen in the model fitting process, consistent with the validation practice for temporally correlated process data [19]. A simulator has provided a full tray profile and steady-state energy balance, and pilot data has calibrated heat loss, condenser approach variation and pressure drop thresholds; hybrid simulation and plant calibration are often employed when complete measurements of the internal column are unavailable [20]. Process historians often have drift and missing values in steam flow tags, and the preprocessing strategy used accepted the dynamic data reconciliation practice [21].

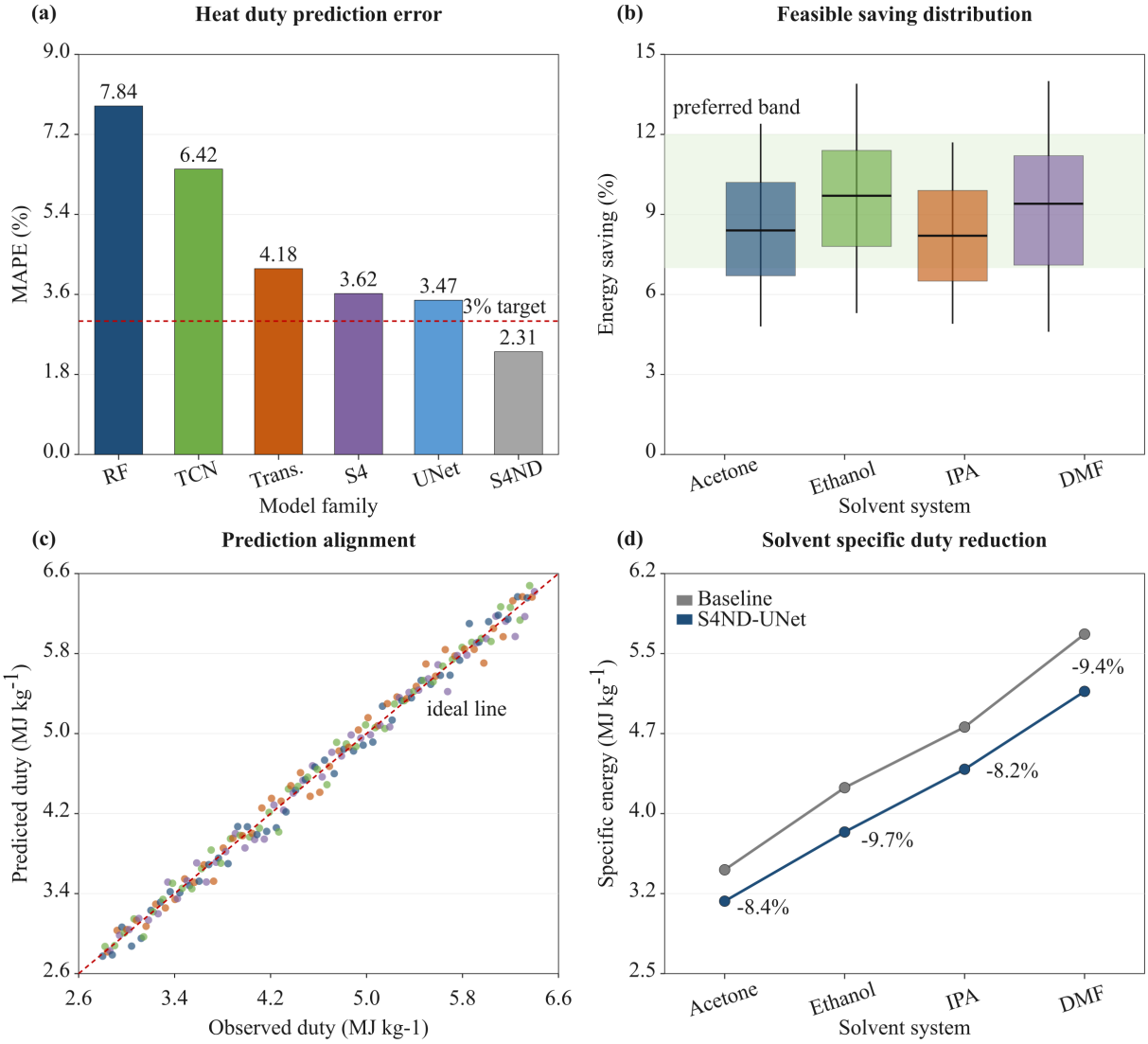


Figure 3. Comparative prediction and energy saving performance: (a) model MAPE, (b) saving distribution, (c) predicted versus observed duty, and (d) solvent specific energy reduction

The four solvent systems are selected to cover all the required ranges of volatility and heat-duty, and they are all subject to the same recovery requirements. Acetone-water was relatively suitable for the separation due to a low specific heat duty; ethanol-water was a typical azeotrope-adjacent recovery case with a narrow purity window for reflux, isopropanol-water exhibited stronger non-ideality, and dimethylformamide-water was a high-boiling solvent that was more prone to heat loss. Change the feed concentration in the different loops during the test, disturb the pressure in the column within the design range, and alter the reflux ratio near the design condition. The resulting windows have normal

operation and feed shock recovery and near-boundary purity control cases. Therefore, to increase the difficulty of the benchmark for the steady-state regression task, it is necessary for the model to learn when to ignore a visible energy reduction due to purity or hydraulic constraints.

Figure 3(a) compares heat duty prediction errors across six model families. The proposed S4ND UNet achieved 2.31% MAPE, compared with 7.84% for a random forest surrogate, 6.42% for temporal convolution, 4.18% for a transformer, 3.62% for a plain S4 model, and 3.47% for a UNet without structured state propagation. The gap of 1.16 percentage points relative to the strongest single mechanism baseline indicates that long memory and spatial profile retention contributed complementary information rather than duplicating one another. Figure 3(b) reports the energy saving distribution under feasible operating search, where the median saving reached 8.9% and the interquartile range remained between 6.7% and 10.8%; the moderate spread shows that the benefit persisted across campaigns instead of being driven by a few unusually favorable windows. Figure 3(c) shows that predicted and observed reboiler duty aligned closely across the 2.8 to 6.4 MJ kg^{-1} range, with the densest region near 4.0 MJ kg^{-1} and only a small dispersion at the high duty boundary, where heat loss uncertainty was strongest. Figure 3(d) separates specific energy by solvent system; ethanol water showed the largest reduction from 4.22 to 3.81 MJ kg^{-1}, while acetone water remained the lowest duty case at 3.17 MJ kg^{-1} after optimization, confirming that percentage saving and absolute duty must be interpreted together.

Based on the comparison, the new gain was not achieved in any of the good-solvent cases. Random Forest Regression can address general non-linearities but fails to capture delayed effects of heat inventory and shows a large error in the disturbance window. Temporal Convolutional Network baseline improved local dynamic prediction but did not have a sufficiently large receptive field for long-term recovery of the trajectory after the feed composition change. The transformer surrogate performed well in the general region; however, it sometimes overestimated a sharp sensor change and proposed a heat reduction that was too large near the purity boundary.

The first S4 and basic UNet models focused on achieving one of the desired behaviors each: the former for long-term memory, and the latter maintained the local profile structure. The full S4ND UNet has combined the above advantages to achieve a lower error distribution for both low- and high-duty-window. Error stratification in different operating modes was introduced. The difference between the Transformer and the full model in the normal window was 0.74 per cent; however, after the feed-shock recovery, it rose to 2.16 per cent. In the upper duty quartile, the median absolute error of the proposed model was less than 0.13 MJ kg^{-1}, and the random forest and temporal convolution baselines exceeded 0.31 MJ kg^{-1}. In this way, the amount of absolute energy has been reduced significantly; however, it is also more susceptible to heat loss and hydraulic uncertainty. Therefore, a lower dispersion of the complete model will increase the number of candidates moves that meet the same conservative acceptance margin without violating purity or pressure-drop limits. Table 1 shows the results of these models at the level of the solvent-specific baseline duty, optimized duty, energy saving, and achieved purity, and also presents prediction accuracy in conjunction with other process performance indicators.

Table 1. Operating benchmark used for energy optimization evaluation

Scenario	Solvent system	Feed range wt%	Samples	Baseline duty MJ kg^{-1}	S4ND UNet duty MJ kg^{-1}	Purity wt%	Energy saving %
S1	Acetone water	14 to 32	4,680	3.46	3.17	99.21	8.38
S2	Ethanol water	18 to 45	4,680	4.22	3.81	99.05	9.72
S3	IPA water	10 to 28	4,680	4.78	4.39	98.94	8.16
S4	DMF water	6 to 19	4,680	5.64	5.11	98.72	9.40
Mean	All systems	6 to 45	18,720	4.53	4.12	98.98	9.04

5.2. Ablation Study and Cross Scale Contribution

The ablation study examined whether the improvement came from structured sequence propagation, cross scale bridges, constrained decoding, or the complete coupling of these components. Removing cross scale bridges increased MAPE from 2.31% to 3.47%, and the mean feasible energy saving decreased from 9.04% to 6.82%. Replacing S4ND blocks with ordinary convolution raised MAPE to 3.91%, with the largest deterioration appearing during slowly changing feed composition. Without the hydraulic feasibility head, the optimizer proposed 11.6% lower heat duty on average, but 14.2% of candidates violated the pressure drop margin, which is not acceptable for plant operation. Constraint aware learning methods have been recommended for process optimization when the surrogate output is used directly for set point selection [22].

Removal of components also changed the type of recommendations. After removing the cross-scale bridges, the average of the predicted duties was still reasonable; however, the optimizer selected reflux reductions that were too large during the short pressure-drop excursion. When S4ND blocks were replaced with a standard convolution, the model

reacted quickly to changes in feed but lost accuracy during the following settling period, and at this time, heat inventory and receiver composition still affected the measured purity margin. Without the constrained decoder, the nominal energy objective was improved, but the number of rejected candidates increased significantly. Therefore, these patterns do not indicate the full extent of the improvement achieved in solvent recovery. Therefore, component-wise ablation must be conducted to determine which architectural modules affect the various engineering results [23].

Figure 4(a) presents the ablation error profile and shows that the complete architecture had the lowest error for duty, purity margin, and hydraulic stress; the simultaneous deterioration of all three outputs after bridge removal indicates that cross scale exchange affected the shared representation rather than one isolated prediction head. Figure 4(b) traces one feed disturbance case in which feed solvent concentration dropped from 31 wt% to 23 wt%; the proposed model stabilized the recommended duty within 18 minutes, while the temporal convolution baseline required 31 minutes, and the smaller post disturbance oscillation reduced repeated reversals of the recommended control move. Figure 4(c) maps the sensitivity of energy saving to reflux ratio and feed concentration, indicating that the best operating pocket occurred around a reflux ratio of 1.46 to 1.58 when feed concentration was between 24 wt% and 36 wt%. Dynamic surrogate studies have reported that multiresolution features can improve disturbance representation when operating regimes change gradually rather than instantaneously [24]. Figure 4(d) compares constraint violation rates, where the full model kept violation below 1.8% across all solvent systems, whereas the unconstrained decoder reached 9.6% for dimethylformamide water; this fivefold difference explains why its apparently larger unconstrained saving cannot be treated as a usable operating gain.

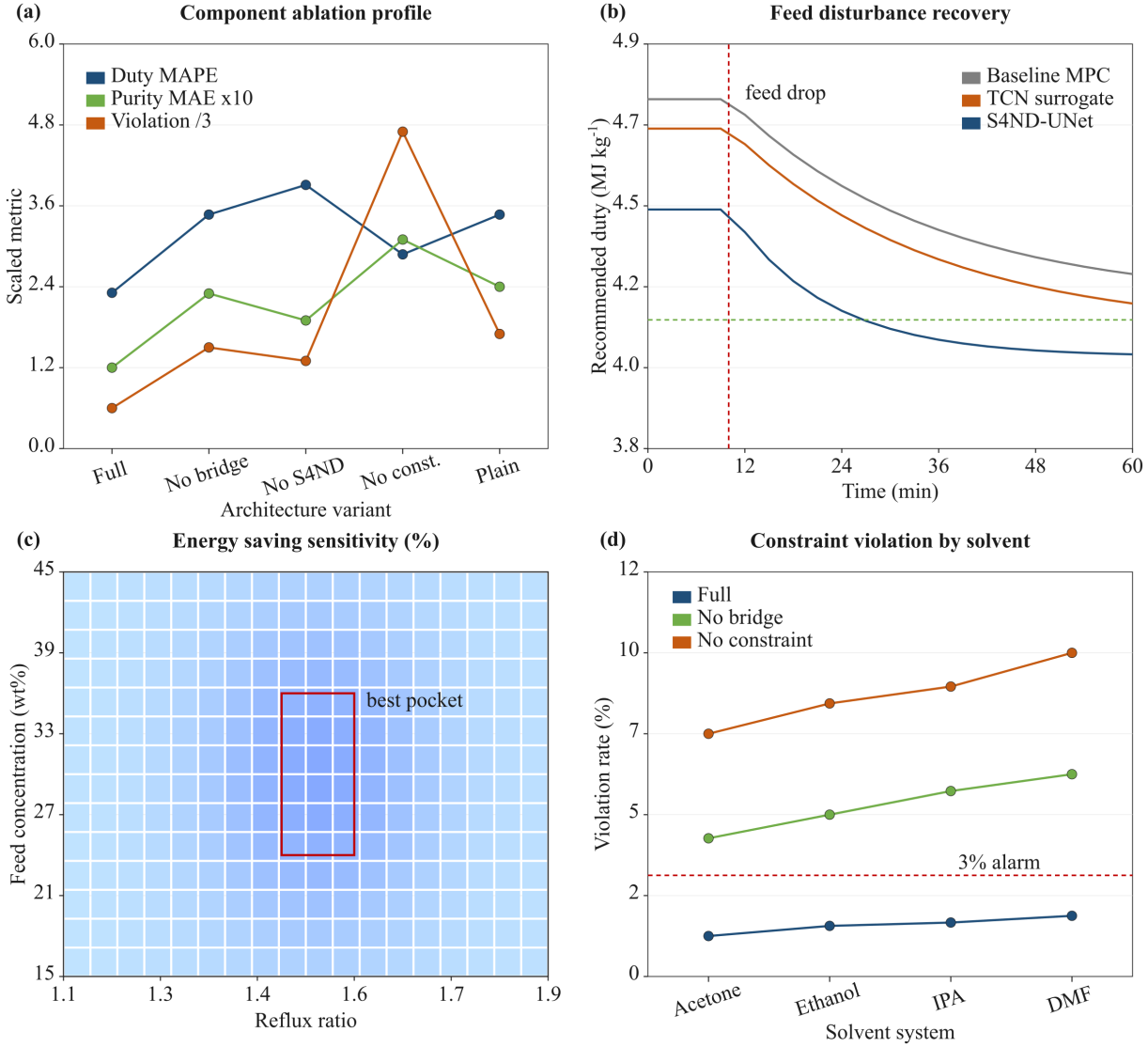


Figure 4. Ablation and disturbance response analysis: (a) component removal errors, (b) duty trajectory under feed disturbance, (c) reflux concentration sensitivity map, and (d) constraint violation rates

The sensitivity map is particularly useful for engineering interpretation because it identifies where the model expects an energy benefit rather than only reporting the final optimum. At reflux ratios below 1.25, the predicted saving is limited by purity risk even though the heat input is lower. At reflux ratios above 1.75, purity is protected, but the energy

penalty grows faster than the marginal recovery benefit. The best pocket therefore sits between the two extremes and shifts slightly with feed concentration. For operators, this map suggests that the model is learning a structured operating region, not simply recommending the lowest reflux value available in the candidate set. The ablation trajectories also reveal different failure modes rather than a uniform loss of accuracy. Removing S4ND primarily increased lag after the disturbance, while removing cross scale bridges increased short duration overshoot near the pressure drop limit. The unconstrained decoder behaved differently again: its energy prediction remained close to the full model, but its candidate ranking promoted points with insufficient purity margin. These distinctions explain why the three components cannot be replaced by a single increase in network width. They address memory, local global feature exchange, and decision feasibility, respectively, and the measured benefit appears when those functions operate together.

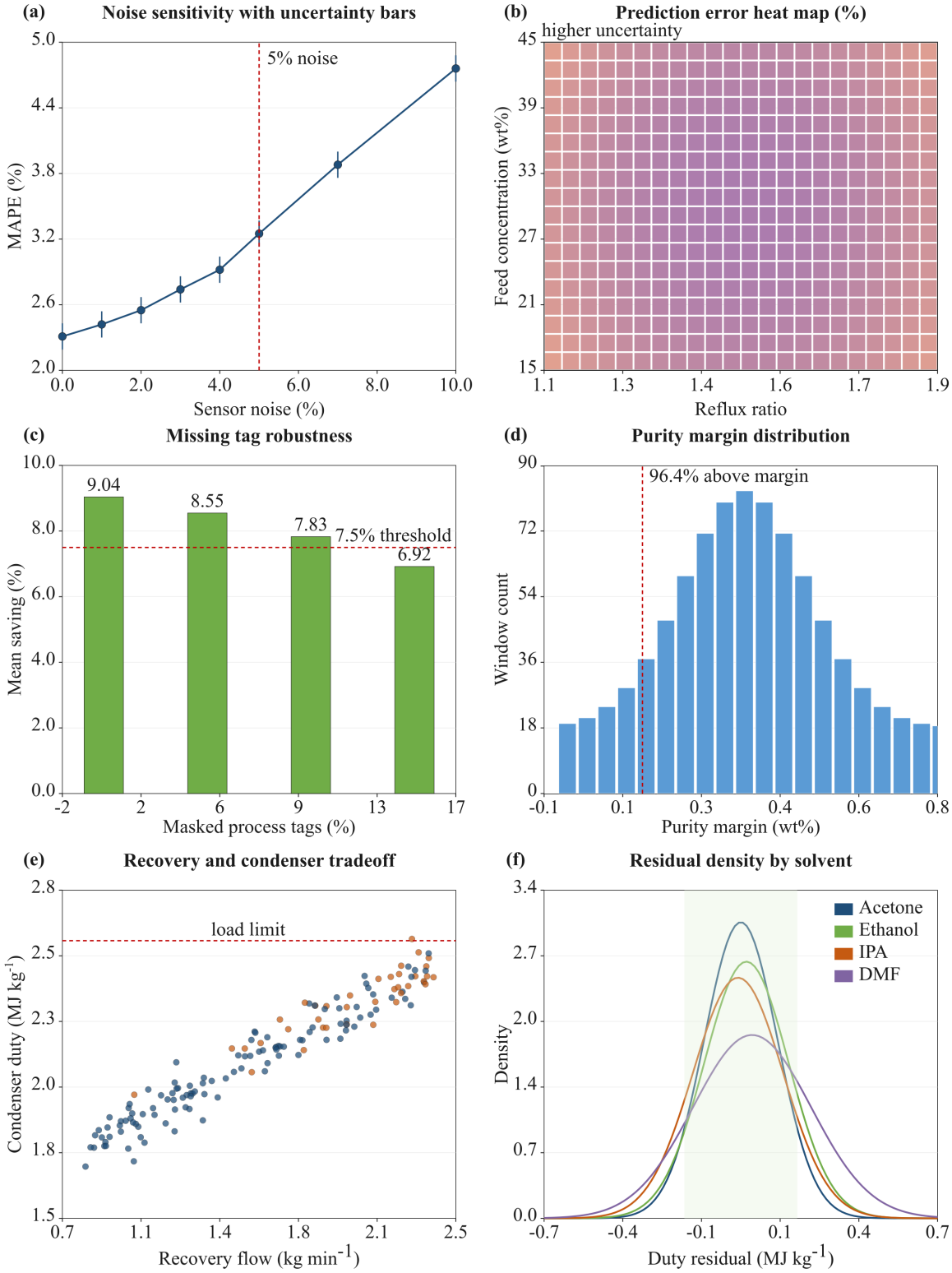


Figure 5. Robustness and uncertainty evaluation: (a) noise sensitivity, (b) operating region error heat map, (c) missing tag energy saving, (d) purity margin distribution, (e) recovery condenser tradeoff, and (f) residual density by solvent system

5.3. Robustness, Generalization, and Operating Interpretation

Robustness was evaluated under sensor noise, missing tags, feed shocks and underrepresented concentration bands. With 3% Gaussian noise added to temperature and flow channels, MAPE rose to 2.74% while mean feasible energy saving fell to 8.31%, indicating more conservative candidate acceptance. When 10% of process tags were masked and reconstructed by interpolation, MAPE reached 3.02%, above the proposed model's clean-data error. In high-feed-concentration windows that were deliberately underrepresented during training, mean saving remained 7.46%. Generalization is important because campaign-specific solvent mixtures do not follow a fixed feed distribution [25].

Uncertainty visualization has been recommended as a companion to aggregate error reporting in data driven process optimization [26]. Figure 5(a) shows the noise robustness curve, where prediction error increased nearly linearly until 5% noise and then steepened as pressure drop uncertainty began to affect hydraulic classification. Figure 5(b) presents the operating region heat map: the lowest error zone occurred at moderate reflux and mid-range feed composition, whereas high concentration combined with low reflux produced a narrow high uncertainty band beside the purity boundary. Figure 5(c) compares energy saving at missing tag ratios from 0% to 15%; saving stayed above 7.5% through 10% masking, although the solvent spread widened. Figure 5(d) plots purity margin, with 96.4% of windows above the 0.15 wt% safety margin; the lower tail therefore governed rejection more strongly than the median. Figure 5(e) shows that selected set points avoided condenser load above 2.6 MJ kg^{-1} without clustering at low recovery flow. Figure 5(f) displays residual density by solvent system; dimethylformamide water had the widest tail, which identifies the mixture for which heat loss calibration and uncertainty screening require the greatest attention.

Figure 6 integrates energy saving, purity margin and hydraulic feasibility for 420 candidates retained from bounded searches across the test campaigns after purity, hydraulic and move-limit screening. Of these candidates, 71% lay in the region with 7% to 12% energy saving and purity margin above 0.20 wt%. Only 1.5% entered the caution band with hydraulic stress above 0.86, and the supervisory rule rejected those points. The operating pocket indicates that duty reduction arose from coordinated reflux, pressure and condenser adjustments rather than uniform boilup reduction. Industrial solvent recovery studies support coordinated set point movement when feed conditions change [27]. Thermodynamic descriptions of non-ideal aqueous mixtures also show that error increases outside the calibrated composition range when activity coefficients vary strongly. Across the four systems, mean specific duty decreased from 4.53 to 4.12 MJ kg^{-1} and solvent-specific energy saving ranged from 8.16% to 9.72%. Mean product purity remained 98.98 wt%, indicating that utility reduction was not obtained by relaxing recovery requirements. Site acceptance testing should therefore retain solvent-specific checks rather than relying only on an aggregate saving value.

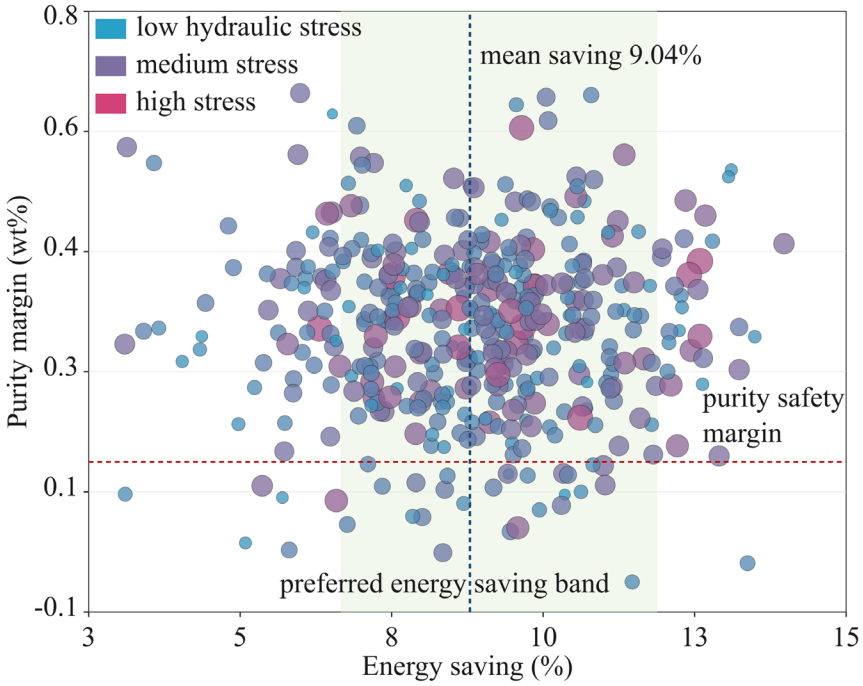


Figure 6. Integrated operating map of energy saving, purity margin, and hydraulic feasibility for feasible candidate selection

Given the limitations of the historical records, cross-scale fusion was employed to stably determine the temperature and condenser channel without a reflux tag. Pressure-drop noise caused a more rapid decline because hydraulic feasibility is directly related to the presence of vapor loading; reduced confidence in this channel appropriately limited aggressive recommendations. Staged operator review before supervisory target adjustment is in line with the established

deployment practice for learning-based process control. Measurement reconciliation research has found that persistent utility meter bias is an independent risk of random sensor noise because interpolation cannot correct for it [28]. Inference is still in advisory mode, only one window is required, and 240 candidates have been excluded before the next supervisory update. The actual bottleneck was stable steam flow correction and condenser duty estimation; therefore, the shadow mode should keep the laboratory purity and utility meter data before the bounded target update. Most of the improvements occurred after changing several mild factors at the same time: a small reduction in reflux was accompanied by an adjustment of the condenser approach and a limited pressure change; otherwise, single large movements of reflux or boil-up were rejected more frequently near the purity and hydraulic limits.

6. Conclusion

This study developed and evaluated a cross-scale S4ND-UNet surrogate for energy optimization in dynamic solvent recovery. The results showed that combining long operating memory, local tray behavior and explicit feasibility outputs improved both prediction and bounded operating selection.

1. The proposed model achieved a heat-duty MAPE of 2.31% and reduced mean reboiler duty from 4.53 to 4.12 MJ kg^{-1}, corresponding to 9.04% mean energy saving while maintaining 98.98 wt% mean product purity.

2. The complete architecture outperformed temporal-convolution and transformer surrogates, whose MAPE values were 6.42% and 4.18%, respectively. Removing cross-scale bridges increased MAPE to 3.47%. In a representative feed disturbance, the complete model stabilized its duty recommendation within 18 min, compared with 31 min for temporal convolution. The constrained decoder kept solvent-specific violation rates below 1.8%, whereas the unconstrained variant reached 9.6% for dimethylformamide-water.

3. The main contribution was the coupling of structured long-range propagation, gated cross-scale exchange and feasibility-aware decoding within one engineering surrogate. Current applicability is limited to bounded advisory moves inside the calibrated envelope and to the four aqueous solvent systems represented in the benchmark. Long-term closed-loop industrial performance remains unverified, and reactive, strongly azeotropic or degradation-sensitive mixtures may require additional thermodynamic descriptors and stricter safety constraints.

The framework provides a basis for energy-aware supervisory optimization. Future work should examine online adaptation, plant-scale heat integration, sensor-bias diagnosis, uncertainty calibration and closed-loop validation under changing solvent campaigns.

7. Declarations

7.1 Author Contributions

Jelena Stojanović contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Ognjenka Ćukić contributes to validation, analysis, investigation, data collection, draft preparation. All authors have read and agreed with the manuscript before its submission and publication.

7.2 Funding and Acknowledgments

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

7.3 Institutional Review Board Statement

Not applicable.

7.4 Informed Consent Statement

Not applicable.

7.5 Data Availability

Not applicable.

7.6 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7.7 Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

8. References

[1] Zhai, J., Xie, H., Chen, X., Peng, Z., Sun, Q., & Li, J. (2024). Design and energy-saving strategy of sustainable pressure-swing distillation with thermally and electrically coupled intensification for separating ternary mixture with multiple azeotropes. Energy, 295, 131078. https://doi.org/10.1016/j.energy.2024.131078

[2] Yin, C., Jiang, G., Wang, Z., Zhang, D., Guo, L., & Liu, G. (2025). Synthesis of distillation sequence with sloppy split and thermally coupled column and its integration with reactor and heat exchanger network. Chemical Engineering Science, 306, 121208. https://doi.org/10.1016/j.ces.2025.121208

[3] Karaman, Ö. F., Lang, P., & Hegely, L. (2025). Optimisation of pressure-swing distillation of a maximum-azeotropic mixture with heat integration. Energy, 320, 135190. https://doi.org/10.1016/j.energy.2025.135190

[4] Wu, L., Lu, K., Li, Q., Xu, L., Luo, Y., & Yuan, X. (2024). Energy-saving design and optimization of pressure-swing-assisted ternary heterogenous azeotropic distillations. Chinese Journal of Chemical Engineering, 68, 1-7. https://doi.org/10.1016/j.cjche.2023.11.026

[5] Wavomba Mtogo, J., Wanyaga Mugo, G., & Mizsey, P. (2024). Enhancing exergy efficiency and environmental sustainability in pressure swing azeotropic distillation. Cleaner Energy Systems, 9, 100134. https://doi.org/10.1016/j.cles.2024.100134

[6] Tang, W. T., Chien, C. K., & Ward, J. D. (2025). A review of energy intensification strategies for distillation processes: Cyclic operation, stacking, heat pumps, side-streams, dividing walls and beyond. Separation and Purification Technology, 357, 130030. https://doi.org/10.1016/j.seppur.2024.130030

[7] Wang, Y., Zhu, W., Cheng, H., Zhong, J., Li, X., Qi, J., Wang, S., Zhu, Z., Wang, Y., & Cui, P. (2024). Enhanced reaction extraction distillation via multi-objective optimization and 4E analysis for sustainable recovery of organic compounds from wastewater. Process Safety and Environmental Protection, 186, 884-894. https://doi.org/10.1016/j.psep.2024.04.064

[8] Niazi, S., Díaz-López, J. A., & Nieto-Márquez, A. (2025). Heat integration and process improvement of a batch distillation unit using a gradual feeding strategy: A case study on MIBK-water separation. Separation and Purification Technology, 362, 131956. https://doi.org/10.1016/j.seppur.2025.131956

[9] Yu, A., Ye, Q., Li, J., Li, X., Wang, Y., & Rui, Q. (2024). Economic, environmental, energy, exergy (4E) analysis and simulated annealing algorithm optimization of dividing-wall column-intensified heterogeneous azeotropic pressure-swing distillation process. Energy, 296, 131099. https://doi.org/10.1016/j.energy.2024.131099

[10] Alizadeh, M. H., & Sahlodin, A. M. (2024). Energy-efficient semi-continuous distillation of a ternary mixture using combined tracking-economic model predictive control. Chemical Engineering Research and Design, 209, 283-298. https://doi.org/10.1016/j.cherd.2024.07.064

[11] Cheng, H., Wang, K., Wang, Y., Zhao, G., Zhu, Z., Qi, J., Zhong, L., Wang, Y., Zhang, H., & Cui, P. (2024). Mechanism Analysis and Multiobjective Optimization of Efficient and Energy-Saving Separation of Green Fuel Additives via Extractive Pressure Swing Distillation with an Ionic Liquid Mixed Entrainer. Industrial & Engineering Chemistry Research, 63(12), 5261-5275. https://doi.org/10.1021/acs.iecr.4c00009

[12] Zhu, Y., Yang, F., & Torgashov, A. (2024). Causal-Transformer: Spatial-temporal causal attention-based transformer for time series prediction. IFAC-PapersOnLine, 58(14), 79-84. https://doi.org/10.1016/j.ifacol.2024.08.317

[13] Bonassi, F., Andersson, C., Mattsson, P., & Schön, T. B. (2024). Structured state-space models are deep Wiener models. IFAC-PapersOnLine, 58(15), 247-252. https://doi.org/10.1016/j.ifacol.2024.08.536

[14] Alhajeri, M. S., Ren, Y. M., Ou, F., Abdullah, F., & Christofides, P. D. (2024). Model predictive control of nonlinear processes using transfer learning-based recurrent neural networks. Chemical Engineering Research and Design, 205, 1-12. https://doi.org/10.1016/j.cherd.2024.03.019

[15] Wan, J., Xia, N., Yin, Y., Pan, X., Hu, J., & Yi, J. (2024). TCDformer: A transformer framework for non-stationary time series forecasting based on trend and change-point detection. Neural Networks, 173, 106196. https://doi.org/10.1016/j.neunet.2024.106196

[16] Pessoa, P., Campitelli, P., Shepherd, D. P., Ozkan, S. B., & Pressé, S. (2025). Mamba time series forecasting with uncertainty quantification. Machine Learning: Science and Technology, 6(3), 035012. https://doi.org/10.1088/2632-2153/adec3b

[17] Zhou, Z., Guo, X., Xiong, Y. J., & Xia, C. M. (2024). Kalman-SSM: Modeling Long-Term Time Series With Kalman Filter Structured State Spaces. IEEE Signal Processing Letters, 31, 2470-2474. https://doi.org/10.1109/lsp.2024.3457862

[18] Hu, G., Xu, L., Zhang, Z., Armaou, A., Hong, Z., & Ni, X. (2025). Dynamic data reconciliation based on multiple models for enhancing the performance of nonlinear dynamic systems. Computers & Chemical Engineering, 202, 109319. https://doi.org/10.1016/j.compchemeng.2025.109319

[19] Zhu, W., Zhu, J., Liu, Y., Zhang, Z., Huang, S., & Zhao, S. (2025). Dynamic Data Reconciliation to Improve the Control Performance of Nonlinear System with Unknown Process Models. Industrial & Engineering Chemistry Research, 64(29), 14545-14559. https://doi.org/10.1021/acs.iecr.5c01048

[20] Zhu, W., Jia, M., Zhang, Z., & Liu, Y. (2024). Dynamic data reconciliation for enhancing the performance of kernel learning soft sensor models considering measurement noise. Chemometrics and Intelligent Laboratory Systems, 246, 105083. https://doi.org/10.1016/j.chemolab.2024.105083

[21] Horak, W., Louw, T. M., & Bradshaw, S. M. (2024). Constrained Multi-Output Gaussian Process Regression for Data Reconciliation. IFAC-PapersOnLine, 58(4), 324-329. https://doi.org/10.1016/j.ifacol.2024.07.238

[22] Peng, Z., Yuan, Y., Chen, Z., Xu, X., & Dubljevic, S. (2025). An EM/ECME-based algorithm for simultaneous data reconciliation and gross error detection using t-distribution noise model in chemical industry. Computers & Chemical Engineering, 202, 109301. https://doi.org/10.1016/j.compchemeng.2025.109301

[23] Guth, S., Mojahed, A., & Sapsis, T. P. (2024). Quality measures for the evaluation of machine learning architectures on the quantification of epistemic and aleatoric uncertainties in complex dynamical systems. Computer Methods in Applied Mechanics and Engineering, 420, 116760. https://doi.org/10.1016/j.cma.2024.116760

[24] Ban, Z., & Pfeiffer, C. (2025). Dynamic data reconciliation with simultaneous time-varying parameter estimation in real time: application to an electric submersible pump lift oil production. Journal of Petroleum Exploration and Production Technology, 15(11), 181. https://doi.org/10.1007/s13202-025-02098-1

[25] Sansana, J., Rendall, R., Castillo, I., Chiang, L., & Reis, M. S. (2024). Hybrid modeling for improved extrapolation and transfer learning in the chemical processing industry. Chemical Engineering Science, 300, 120568. https://doi.org/10.1016/j.ces.2024.120568

[26] Durkin, A., Otte, L., & Guo, M. (2024). Surrogate-based optimisation of process systems to recover resources from wastewater. Computers & Chemical Engineering, 182, 108584. https://doi.org/10.1016/j.compchemeng.2024.108584

[27] Chen, Y., Zhang, D., Yan, R., & Xie, M. (2025). Applications of Domain Generalization to Machine Fault Diagnosis: A Survey. IEEE/CAA Journal of Automatica Sinica, 12(10), 1963-1984. https://doi.org/10.1109/jas.2025.125120

[28] Chai, Z., Zhao, C., & Huang, B. (2025). Cross-domain knowledge transfer in industrial process monitoring: A survey. Journal of Process Control, 149, 103408. https://doi.org/10.1016/j.jprocont.2025.103408