Received: 07 September 2026; **Revised:** 01 November 2026; **Accepted:** 03 November 2026; **Published:** 20 November 2026
Archs Sci. (2026) Volume 76, Issue 5 Pages 233-242, Paper ID 2026524.
https://doi.org/10.68304/as/76524

Forecast-Driven Trading Optimization for Electricity Retailers Under Demand Response Uncertainty and Asymmetric Imbalance Settlement

Shasha Zhao[1,*] and Yakui Zhu[1]
[1]State Grid Hebei Electric Power Co., Ltd. Marketing Service Center, Shijiazhuang, Hebei 050000, China
Correspondence: zhaoshasha520@126.com

Abstract To address the complex interdependencies among demand forecasting errors, uncertainties in user response delivery, and asymmetric deviation settlement in multimarket electricity trading, this paper proposes a forecast-driven, risk-aware optimization framework for electricity procurement and sales. A CNN–LSTM model is employed to extract local load features and temporal dependencies, and produces demand forecasts for the following 24 hours. By integrating both price-based and incentive-based demand response, the framework characterizes user electricity consumption adjustments and the uncertainty in actual response delivery. The study formulates a two-stage distributionally robust optimization model to coordinate electricity procurement, retail pricing, flexibility resource allocation, and real-time corrective actions, with Conditional Value-at-Risk (CVaR) used to quantify tail risks. Based on hourly load data, the study conducts forecast performance comparisons and operational scenario analyses. The results demonstrate that the integrated model follows the observed fluctuations in electricity demand; furthermore, price-based responses and energy storage scheduling can reshape the temporal distribution of electricity procurement and consumption—though peak shaving and price reductions do not necessarily lead to a decrease in trading deviations.

Index Terms electricity retail company; load forecasting; demand response; electricity procurement and sales optimization; distributionally robust optimization

I. Introduction

The development of electricity market liberalization has transformed electricity retailing from relatively straightforward procurement and sales into an activity requiring comprehensive decision-making that encompasses multi-time-scale trading, differentiated retail contracts, and the management of demand-side flexibility [1]. Electricity retailers must secure base-load electricity through medium- to long-term contracts, participate in day-ahead markets based on demand forecasts, and—during the actual operational phase—bear the real-time settlement risks arising from imbalances between supply and demand [2], [3]. Meanwhile, end-user electricity consumption is influenced by factors such as weather conditions, production schedules, behavioral patterns, and price incentives, rendering it cyclical, non-stationary, and stochastic. Any deviation between procurement plans and actual demand not only alters procurement costs but can also amplify operational losses, particularly when market prices are high. Consequently, a central challenge for electricity retailers is to coordinate demand forecasting, user response, and multi-market trading—balancing economic returns with risk control while ensuring that electricity demand is met and operational constraints are satisfied [4].

Demand response provides flexible resources that can help address these operational and market challenges. Through mechanisms such as time-of-use pricing, dynamic pricing, and incentive-based compensation, electricity retailers can guide users to shift or curtail a portion of their electricity demand, thereby reducing reliance on high-cost real-time power procurement [5]. Energy storage systems, by enabling the storage and discharge of energy across different time periods, offer further control over the timing of external power procurement. However, the economic value of these flexible resources is not determined solely by their adjustable capacity. User-side adjustments entail opportunity costs; discrepancies may arise between committed and actual response; and the operation of energy storage involves efficiency losses and degradation costs. Furthermore, simultaneous responses to price signals by many users may create new demand peaks. Consequently, an effective trading strategy must integrate user utility, resource delivery capabilities, and market settlement mechanisms into a unified analytical framework, rather than focusing exclusively on objectives such as reducing peak demand or eliminating deviations in electricity consumption [6].

Existing research on the trading decisions of electricity retailers has primarily focused on the configuration of electricity procurement and sales portfolios, as well as the measurement of associated risks. For instance, Sun et al. developed a trading portfolio optimization model for electricity retailers operating within the Chinese power market; they employed Conditional Value-at-Risk (CVaR) to quantify procurement and sales risks and analyzed how different risk preferences influence trading configurations [1]. Such research incorporates both financial returns and exposure to risk into the decision-making process, providing a foundation for multi-market operations. However, in the context of the issues addressed in this paper, trading portfolio optimization must be further integrated with specific consumer adjustment behaviors—specifically, by clarifying how demand response alters actual post-delivery loads and the subsequent outcomes of deviation settlement.

Another line of research focuses on the interplay between retail pricing, user response, and market transactions. Qiu et al. [2] propose a data-driven model for retail pricing and demand declaration, employing a CNN–LSTM architecture to extract both local load features and long-term dependencies. It further addresses the relationship between demand response and market clearing through a bilevel optimization framework incorporating chance constraints. This indicates that integrating load forecasting with transaction optimization is already well-grounded in existing literature; therefore, merely employing a hybrid neural network or linking the forecasting process with the pricing mechanism is insufficient to constitute a novel methodological contribution. This paper should instead focus on how forecasting errors, the delivery of demand responses, and two-way settlement collectively influence transaction decisions—with a particular emphasis on distinguishing between improvements in forecasting accuracy and gains in economic efficiency.

Addressing the challenge of accurately characterizing user heterogeneity and response relationships, Che et al. [3] propose an integrated learning and optimization-based pricing approach for residential and industrial users. This method employs distinct frameworks to model the demand response of these two user categories, utilizing historical price and electricity consumption data to learn the response behavior of industrial users. Such an approach offers a valuable reference for pricing decisions in scenarios where production-related information is not fully transparent; it also highlights that user response cannot be simply equated to a pre-specified, dispatchable capacity. From a risk-management perspective, beyond the average price-response relationship, it is essential to account for fluctuations in the actual delivery ratio, as well as the potential for correlated deviations among multiple users driven by shared external factors [7].

The coordination of bidding and incentives within demand response markets has also garnered attention. Huang et al. [8] establish a two-stage joint bidding and pricing framework for load aggregators, incorporating uncertain user behavior, risk quantification, and user choice into the decision-making process; it highlights the close linkage between bid quantities and incentive prices. Consequently, joint consideration of response uncertainty and bid pricing has therefore been addressed in prior work. This paper focuses specifically on the procurement and sales scenarios of electricity retailers. It investigates how—in the presence of both demand forecasting errors and response delivery deviations—to determine the scale of flexibility resource deployment and the allocation of responsibility for any residual deviations, while accounting for asymmetric real-time settlement mechanisms [9].

Based on the research discussed above, this paper focuses on three interrelated challenges. First, forecasting objectives do not perfectly align with trading objectives. A model with a low average forecast error may still underestimate demand during periods of high prices, leading to significant real-time procurement costs; therefore, the evaluation of forecasting models must account for the magnitude of errors, the timing of their occurrence, and their associated economic consequences. Second, the bid price of a flexible resource is not equivalent to its actual delivery cost. Resources with low bids but unreliable delivery may incur additional settlement expenses due to insufficient response; conversely, resources with higher prices but reliable delivery may not necessarily be optimal to dispatch in every time period. Finally, demand forecast errors and resource responses across multiple time periods are often interdependent. Constructing scenarios on an hour-by-hour, independent basis can distort load trajectories and resource constraints, thereby compromising electricity procurement strategies, energy storage management, and the assessment of tail risks.

To address these issues, this paper develops a prediction-driven optimization framework for risk-aware electricity procurement and sales by electricity retailers. The prediction layer employs a CNN–LSTM model to generate baseline demand forecasts for the next 24 hours, deriving error samples through rolling forecasting. The user layer incorporates both price-based and incentive-based responses to characterize demand adjustments and actual delivery. The trading layer distinguishes between ex-ante decisions—covering electricity procurement, pricing, and flexibility resource allocation—and real-time corrective actions; it optimizes procurement costs, retail revenue, response compensation, and deviation settlement, subject to constraints on information availability. Furthermore, a data-driven distributional ambiguity set is utilized to characterize joint uncertainty, with Conditional Value-at-Risk (CVaR) employed to quantify potential losses in adverse scenarios. In an extended scenario involving energy storage, the storage system is modeled as an inter-temporal adjustment resource, subject to constraints on capacity, power, efficiency, and terminal energy levels. The formulation permits residual deviations when the cost of demand response exceeds the cost of managing those deviations, thereby establishing optimal adjustment boundaries from both economic and risk-management perspectives.

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

II. Proposed Approach

A. Multi-Time-Scale Electricity Procurement and Sales and Two-Way Demand Response Trading Model

Figure 1 illustrates the transaction structure through which an electricity retailer connects the wholesale market with end-users. In this diagram, the solid blue lines represent the flow of electricity, while the dashed red lines represent the flow of financial settlements. On the procurement side, medium- to long-term contracts secure a baseline supply, day-ahead trading refines procurement in response to updated forecasts, and real-time transactions settle the remaining differences between procured electricity and actual demand. On the retail side, the retailer offers fixed-tariff, time-of-use, and real-time-tariff contracts, with upper and lower price bounds applying to the latter. These contract structures allocate price exposure differently, allowing retailers to develop portfolios that reflect customers' electricity consumption profiles and their tolerance of price uncertainty.

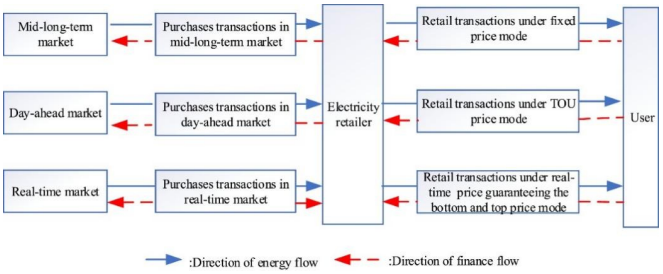


Figure 1: Multi-timescale electricity procurement and sales structure

Building upon the existing interruptible load scheme, the scope of user-side adjustment is expanded to encompass both load reduction and load increase. Load reduction serves to mitigate situations where actual demand exceeds the volume of electricity already procured, while load increase utilizes verifiable resources—such as shiftable production tasks or controllable charging—to absorb surplus electricity. Additional demand must correspond to a genuine service, production activity, or charging need and must comply with equipment limits, production schedules, and user-utility constraints; otherwise, an artificial increase in consumption would not represent a legitimate source of economic value [10], [11].

Let the scheduling horizon comprise $T = 24$ time periods, where $\mathcal{I}$ denotes the set of users and ω representing an uncertain scenario. Let q_t^{ML} be the power allocated to each time period under medium-to-long-term contracts, q_t^{DA} be the power purchased in the day-ahead market, and $q_{t\omega}^{+}$ and $q_{t\omega}^{-}$ be the power purchased or sold, respectively, in the real-time market. The net expenditure for electricity procurement and real-time transactions can then be expressed as:

$$C_{\omega}^{\mathrm{E}} = \sum_{t=1}^{T} \Delta t \left[\lambda_t^{\mathrm{ML}} q_t^{\mathrm{ML}} + \lambda_t^{\mathrm{DA}} q_t^{\mathrm{DA}} + \lambda_{t\omega}^{\mathrm{RT},+} q_{t\omega}^{+} - \lambda_{t\omega}^{\mathrm{RT},-} q_{t\omega}^{-}\right], \tag{1}$$

where Δt denotes the duration of the time interval. Since trading volumes are expressed in power units (e.g., MW), multiplying them by Δt yields the settled energy quantity (e.g., MWh). Distinct real-time settlement prices are set for positive and negative volumes, thereby accounting for the asymmetry between purchasing additional energy (to cover a deficit) and disposing of surplus energy. If the market under study does not permit retail electricity companies to sell surplus energy directly in the real-time market, the final term should be replaced by a settlement term for surplus energy that aligns with the specific market rules.

Let k_i denote the type of contract adopted by user i; the retail price is defined as follows:

$$p_{it} = \begin{cases} p_i^{\mathrm{F}}, & k_i = \mathrm{F}, \\ \sum_{b \in B} p_{ib}^{\mathrm{TOU}} \mathbf{1}_{\{t \in T_b\}}, & k_i = \mathrm{TOU}, \\ \min\{\overline{p}_i, \max[\underline{p}_i, a_i + b_i \lambda_t^{\mathrm{ref}}]\}, & k_i = \mathrm{RTP}, \end{cases} \tag{2}$$

where T_b denotes the set of time periods, such as peak, off-peak, and valley periods; $\underline{p}_i$ and $\overline{p}_i$ represent the price protection boundaries; and λ_t^{ref} is the reference price known at the time of pricing. Pre-announced retail prices are therefore linked only to prices observed or forecast before the pricing decision; realized real-time market prices must not enter an earlier decision as if they were known with certainty.

Let the user's price-responsive demand be denoted by $x_{it}(p_i)$. Accounting for prediction errors and the delivery of incentive-based responses, the actual demand and the transaction balance satisfy the following:

$$d_{it\omega} = x_{it}(p_i) + \varepsilon_{it\omega} - \alpha_{it\omega}^{\mathrm{IL}} u_{it}^{\mathrm{IL}} + \alpha_{it\omega}^{\mathrm{UP}} u_{it}^{\mathrm{UP}}, \tag{3}$$

$$q_t^{\mathrm{ML}} + q_t^{\mathrm{DA}} + q_{t\omega}^{+} - q_{t\omega}^{-} = \sum_{i \in \mathcal{I}} d_{it\omega}, \tag{4}$$

$$0 \le q_{t\omega}^{+} \le M_t z_{t\omega}, \quad 0 \le q_{t\omega}^{-} \le M_t (1 - z_{t\omega}), \quad z_{t\omega} \in \{0,1\}, \tag{5}$$

where u^{IL} and u^{UP} represent the agreed-upon power reduction and increase, respectively; $\alpha \in [0,1]$ denotes the actual delivery ratio; and ε represents the demand error. Binary variables are employed to prevent simultaneous purchasing and selling during the same time interval, while M_t should be set based on a physically justified upper bound for transaction capacity.

To avoid treating user response simply as an arbitrary callable quantity, price response is determined by the maximization of user utility:

$$\mathbf{x}_i(\mathbf{p}_i) \in \arg\max_{\mathbf{x}_i \in \mathcal{X}_i} \sum_{t=1}^{T} \Delta t \left[a_{it}^u x_{it} - \frac{b_i^u}{2} x_{it}^2 - p_{it} x_{it} - \frac{\kappa_i}{2} (x_{it} - \hat{d}_{it})^2 \right], \tag{6}$$

$$\mathcal{X}_i = \left\{ \mathbf{x}_i : \underline{x}_{it} \le x_{it} \le \overline{x}_{it},\ E_i \le \sum_t x_{it} \Delta t \le \overline{E}_i,\ |x_{it} - x_{i,t-1}| \le R_i \right\}, \tag{7}$$

where $b_i^u > 0$, $\hat{d}_{it}$ represents the baseline demand forecast, κ_i denotes the cost of deviating from the original electricity consumption plan, and R_i signifies the limit on the rate of load change. This model establishes a clear link between

retail pricing, user utility, and actual load, while providing a user-side response relationship for subsequent bilevel trading optimization.

B. Day-Ahead Load Forecasting Model Based on CNN–LSTM Feature Fusion

Figure 2 illustrates the structure of the demand forecasting module. The model receives a historical electricity consumption sequence spanning H consecutive hours before the forecast origin and produces a 24-hour sequence of expected baseline demand. A convolutional module is employed to extract short-term variations, local peaks and troughs, and recurring load patterns, while an LSTM module is used to capture inter-temporal dependencies. Features from the two branches are then combined and passed through fully connected layers to produce a multi-step forecast, which provides a common demand baseline for day-ahead procurement declarations and retail pricing decisions.

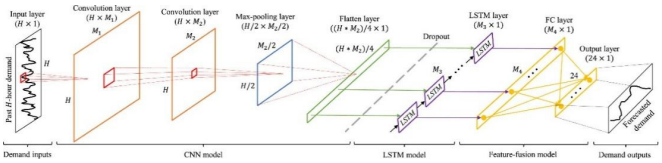


Figure 2: CNN–LSTM feature fusion network for 24-hour-ahead demand forecasting

To ensure consistency between the model description and the provided diagrams, the vectorized features obtained from the convolutional branch can be fused with features extracted independently by an LSTM that retains the temporal order of the input. A single flattened vector without an explicit temporal axis cannot be supplied directly to an LSTM as though it were a sequence; if the architecture instead places the LSTM after the convolutional stages, the convolutional output must first be reshaped into a sequence with a well-defined time dimension [12].

Let the normalized historical input be denoted by $\mathbf{X} \in \mathbb{R}^{H\times F}$. When only historical load is supplied, as depicted in the diagram, $F = 1$; if additional variables—such as weather or calendar data—are incorporated, both the diagram and the experimental setup should be adjusted accordingly. The 1D convolution at layer l is expressed as follows:

$$Z_{h,m}^{(l)} = \phi\left(b_m^{(l)} + \sum_{c=1}^{M_{l-1}} \sum_{r=0}^{K_l-1} W_{r,c,m}^{(l)} Z_{h+r-p_l,c}^{(l-1)}\right), \quad \mathbf{Z}^{(0)} = \mathbf{X}, \tag{8}$$

where K_l , p_l , and M_l denote the kernel length, padding length, and number of output channels, respectively, while ϕ represents a non-linear activation function. Successive convolutional layers extract local load patterns at multiple levels of abstraction from the historical time series.

Pooling and flattening operations yield a compact representation of the convolutional branch:

$$P_{j,m} = \max_{0\le r<s} Z_{sj+r,m}^{(L)}, \tag{9}$$

$$\mathbf{f}^{\mathrm{C}} = \mathrm{vec}(\mathbf{P}), \quad \mathbf{f}^{\mathrm{C}} = \frac{\mathbf{m}\odot\mathbf{f}^{\mathrm{C}}}{1-\rho}, \tag{10}$$

$$m_j \sim \mathrm{Bernoulli}(1-\rho). \tag{11}$$

Here, s denotes both the pooling-window width and the stride, while ρ denotes the dropout probability during training. Dropout is disabled during inference, so predictions are produced without random feature suppression. Pooling reduces the dimensionality of the convolutional feature representation, whereas dropout reduces the tendency of the network to overfit particular historical observations.

The LSTM branch processes the input in chronological order, with its gating and hidden-state updates defined as follows:

$$\mathbf{f}_h = \sigma(\mathbf{W}_f\mathbf{X}_h + \mathbf{U}_f\mathbf{h}_{h-1} + \mathbf{b}_f), \tag{12}$$

$$\mathbf{i}_h = \sigma(\mathbf{W}_i\mathbf{X}_h + \mathbf{U}_i\mathbf{h}_{h-1} + \mathbf{b}_i), \tag{13}$$

$$\mathbf{o}_h = \sigma(\mathbf{W}_o\mathbf{X}_h + \mathbf{U}_o\mathbf{h}_{h-1} + \mathbf{b}_o), \tag{14}$$

$$\mathbf{c}_h = \tanh(\mathbf{W}_c\mathbf{X}_h + \mathbf{U}_c\mathbf{h}_{h-1} + \mathbf{b}_c), \tag{15}$$

$$\mathbf{c}_h = \mathbf{f}_h\odot\mathbf{c}_{h-1} + \mathbf{i}_h\odot\mathbf{c}_h, \quad \mathbf{h}_h = \mathbf{o}_h\odot\tanh(\mathbf{c}_h). \tag{16}$$

The forget, input, and output gates jointly govern the retention, updating, and transmission of information through time, enabling the network to learn dependencies between intraday cycles and consecutive time periods. Once the features from the convolutional layers and the LSTM are fused, the model directly outputs predictions for 24 time periods.

$$\hat{\mathbf{d}}_{n,1:24} = \mathrm{softplus}\big[\mathbf{W}_2\phi\big(\mathbf{W}_1[\mathbf{f}_n^{\mathrm{C}};\mathbf{h}_{n,H}] + \mathbf{b}_1\big) + \mathbf{b}_2\big], \tag{17}$$

$$\min_{\Theta} L(\Theta) = \frac{1}{24N}\sum_{n=1}^{N}\sum_{t=1}^{24} w_t(d_{n,t} - \hat{d}_{n,t})^2 + \lambda_\ell\|\Theta\|_2^2, \tag{18}$$

where the softplus function is used to constrain the predicted load to be non-negative, while w_t represents time-period weights that are either pre-specified or determined solely based on the training set. The data for model training, validation, and testing must be partitioned chronologically, and normalization parameters must be estimated exclusively from the training data. Furthermore, the prediction residuals are utilized for uncertainty modeling in the subsequent section; the predicted values themselves should not be treated directly as the actual demand.

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

C. Joint Optimization of Predictive Demand Response and Market Trading

Figure 3 illustrates the interconnections among historical data, demand forecasting, day-ahead trading, user response, and real-time settlement. Electricity retailers first generate a baseline demand forecast for the next 24 hours using a CNN–LSTM model, which then informs the determination of day-ahead electricity procurement volumes and retail prices. Users adjust their consumption plans based on these electricity prices, while actual delivery is further influenced by factors such as production requirements and equipment availability. Consequently, the forecast results serve as a baseline for decision-making, with actual real-time deviations emerging from a combination of demand forecasting errors, price-based responses, and responses to incentive mechanisms.

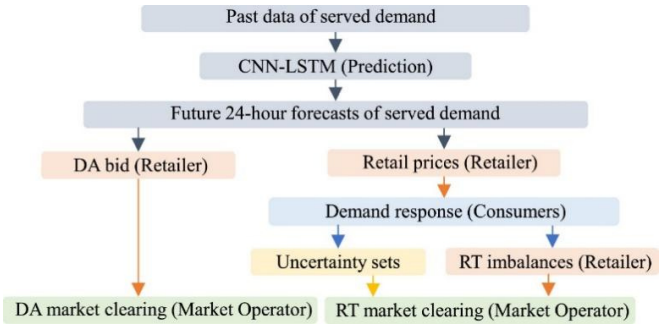


Figure 3: Prediction-driven joint decision-making process for electricity procurement and sales

This process should adhere to a clear informational sequence: day-ahead decisions may only utilize information available prior to the submission deadline, whereas real-time corrective decisions may incorporate realized demand and market data. The day-ahead and real-time clearing processes depicted in the figure are performed by the market operator. Because this study does not formulate system-level market clearing explicitly, wholesale clearing prices enter the optimization as exogenous variables and the retailer is modeled as a price taker [13].

To simultaneously account for demand forecast errors, electricity price volatility, and uncertainties in response delivery, we define a joint random vector along with an ambiguity set for its probability distribution:

$$\boldsymbol{\xi} = (\varepsilon, \boldsymbol{\lambda}^{\mathrm{RT},+}, \boldsymbol{\lambda}^{\mathrm{RT},-}, \boldsymbol{\alpha}^{\mathrm{IL}}, \boldsymbol{\alpha}^{\mathrm{UP}}). \tag{19}$$

Joint samples are intended to preserve dependencies among demand errors, electricity prices, and realized response delivery, with consistent scaling across the different quantities. Where delivery ratios depend on the offered incentive, a conditional response distribution is required: an unconditional distribution calibrated at one incentive level cannot be assumed unchanged after that incentive is modified.

Let $\mathbf{y}$ denote the set of ex-ante decisions, which includes day-ahead electricity procurement, adjustable retail contract parameters, and response reservation quantities; let $\mathbf{r}(\boldsymbol{\xi})$ denote the real-time recourse strategy. The risk-aware joint optimization problem is formulated as follows:

$$\max_{\mathbf{y}\in\mathcal{Y},\ \mathbf{r}\in\mathcal{R}_{\mathrm{NA}}(\mathbf{y})} \inf_{\mathbb{P}\in\mathcal{P}_{\mathbf{y}}} \left\{ \mathbb{E}_{\mathbb{P}}[\Pi(\mathbf{y},\mathbf{r},\boldsymbol{\xi})] - \beta\, \mathrm{CVaR}_{\gamma}^{\mathbb{P}}[-\Pi(\mathbf{y},\mathbf{r},\boldsymbol{\xi})] \right\}, \tag{20}$$

where $\mathcal{R}_{\mathrm{NA}}$ denotes the set of remedial strategies that satisfy the constraints of information availability and non-anticipativity; $\beta \geq 0$ is the risk-aversion coefficient, and γ is the confidence level for tail risk. The objective balances expected retailer profit against exposure to severe losses and does not impose an artificial requirement that every trading imbalance be eliminated.

Scenario profit is calculated by subtracting transaction, response, and other actual costs from retail revenue:

$$\begin{aligned}\Pi_{\omega} = \sum_{i,t} p_{it} d_{it\omega} \Delta t - C_{\omega}^{\mathrm{E}} - \sum_{i,t} \Delta t \big[c_{it}^{\mathrm{IL}} \alpha_{it\omega}^{\mathrm{IL}} u_{it}^{\mathrm{IL}} + c_{it}^{\mathrm{UP}} \alpha_{it\omega}^{\mathrm{UP}} u_{it}^{\mathrm{UP}} \big] \\ - C^{\mathrm{res}}(\mathbf{y}) - C_{\omega}^{\mathrm{op}} - \sum_{t} \Delta t \big[\kappa_t^{+} (q_{t\omega}^{+} - \overline{q}_t^{+})_{+} + \kappa_t^{-} (q_{t\omega}^{-} - \overline{q}_t^{-})_{+} \big],\end{aligned} \tag{21}$$

where C^{res} represents the reservation cost for response capacity, C^{op} denotes other operational costs, and $(a)_{+} = \max(a, 0)$. The final term is relevant only when market rules impose an additional imbalance assessment. If this charge has already been incorporated into real-time settlement prices, it should be omitted to avoid counting the same cost twice. Compensation payments are settled based on the actual delivered volume, with the specific payment method determined in accordance with the contract.

Let scenario loss be $L_{\omega} = -\Pi_{\omega}$. For a discrete scenario distribution, CVaR can be expressed as follows:

$$\mathrm{CVaR}_{\gamma}^{\mathbb{P}}(L) = \min_{\eta, \{s_{\omega}\}} \left[\eta + \frac{1}{1-\gamma} \sum_{\omega\in\Omega} \pi_{\omega} s_{\omega} \right], \tag{22}$$

$$s_{\omega} \geq -\Pi_{\omega} - \eta, \quad s_{\omega} \geq 0, \quad \sum_{\omega\in\Omega} \pi_{\omega} = 1. \tag{23}$$

This representation introduces tail losses in a form amenable to numerical optimization. Whether the complete formulation admits an equivalent mixed-integer linear reformulation nevertheless depends on the treatment of price–demand products, the user-level optimization problem, and the distributional ambiguity set; the precise reformulation assumptions and numerical solution procedure must therefore be specified.

III. Experimental Design and Results Analysis

This section analyzes the prediction-driven electricity procurement and sales optimization scheme proposed in this paper, focusing on the interplay between demand forecasting, user demand response, energy storage dispatch, and market trading. The analysis begins by describing the experimental data and its statistical characteristics, followed by a comparison of load forecasting performance across different models [14]. It then examines user electricity consumption adjustments and energy storage operations in response to price signals. Finally, it investigates how demand declarations, market

prices, and trading deviations evolve under various parameter scenarios [15].

A. Experimental Data and Preprocessing

This study utilizes hourly service load sequences to establish a data foundation for demand forecasting and the optimization of electricity procurement and sales. The temporal variations and frequency distributions of these data are presented in Figures 4 and 5, respectively. The dataset consists of time-stamped electricity-load observations, with the load measured in megawatts (MW). As shown in the figures, the sequence spans the period from early 2018 to early 2020, allowing for the examination of demand fluctuations across different seasons and load levels. The source of these measurements, exact start and end dates, number of valid observations, and electricity-market region remain insufficiently documented in the supplied material and require confirmation against the original data records.

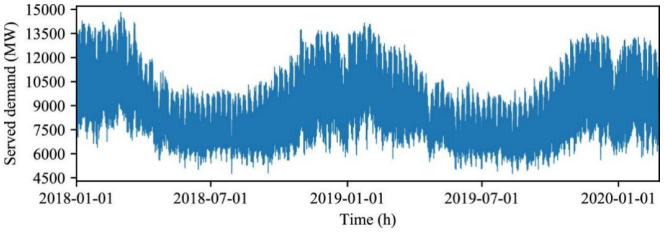


Figure 4: Hourly service load time series

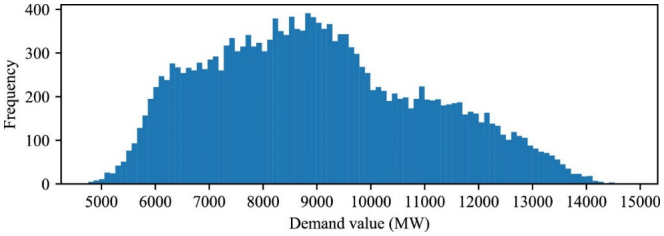


Figure 5: Frequency distribution of service loads on an hourly basis

Figure 4 illustrates the variation in service load over time, characterized by distinct seasonal fluctuations superimposed with frequent, short-term oscillations. The curves show that load levels are generally higher at the beginning and end of the year and relatively lower toward the middle, with consistent patterns observed across different years. At the same time, the magnitudes of local peaks and troughs vary, indicating that load dynamics are both repetitive and non-stationary. The plotted load appears to fall generally within the 5,000–15,000 MW range, although the precise minimum and maximum require calculation from the underlying records.

These characteristics align with the CNN–LSTM forecasting architecture employed in this study: the convolutional module extracts local variations within adjacent hours, while the LSTM module captures dependencies across consecutive time periods. The integrated features are then used to generate demand forecasts for the next 24 hours. In the context of trading models, the presence of seasonal and short-term fluctuations implies that a fixed power procurement ratio cannot consistently match actual demand. Consequently, beyond medium- to long-term baseline procurement, it is necessary to adjust forecasted demand through day-ahead trading, while addressing any remaining deviations via real-time settlement and demand response mechanisms. However, whether the observed short-term oscillations correspond specifically to daily or weekly cycles requires further verification through autocorrelation analysis or periodic decomposition.

Figure 5 further illustrates the frequency of occurrence for various load levels. The load distribution spans a wide range, with a primary frequency peak situated between approximately 8,000 and 9,500 MW. It extends toward higher load levels, exhibiting a "shoulder" feature in the 11,000–12,000 MW range. This indicates that the sample is not uniformly distributed and that the frequency of occurrence varies across different operating states. While this pattern may be linked to a combination of seasonal factors and specific times of day, it is insufficient to conclude that the data follows a normal distribution or any specific mixture distribution.

Given these distributional characteristics, the experiments must account for both typical and high-load periods. Relying solely on an aggregate average prediction error may mask a model's shortcomings during periods of high demand—instances where underestimating demand could necessitate additional real-time procurement. Consequently, the experiments in this study evaluate prediction errors across the entire dataset, during peak periods, and for different seasons. Furthermore, the assessment of trading performance includes an examination of net profit, the volume of real-time procurement, deviation settlement costs, and tail risk. It should be noted that Figure 5 depicts the distribution of the raw load data; it cannot serve as a direct proxy for the distribution of prediction residuals, nor can it, on its own, quantify the magnitude of price risk or the risk associated with demand-response performance [16].

To preserve the temporal causality of the experiments, the data are partitioned into training, validation, and test sets in chronological order. The training set is used to learn the CNN–LSTM parameters; the validation set is employed to select the historical window length, network hyperparameters, and risk model parameters; and the test set is used to evaluate out-of-sample performance. Input samples consist of observations from the H consecutive hours preceding the prediction time, while the corresponding output represents the load for the subsequent 24 hours. Normalization parameters are calculated exclusively from the training set, and the handling of missing values or outliers is performed without incorporating information from periods following the prediction time. Rolling-origin forecasts are evaluated during the testing phase, and various electricity procurement and sales strategies are compared under identical information conditions.

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

B. Reference Models and Operational Scenarios

For load forecasting, four model families are compared: ANN, CNN, LSTM, and CNN–LSTM. Each model utilizes the same data partitioning, input information, and forecasting horizon. Following hyperparameter tuning using the validation observations for each model, the models are compared to evaluate the effectiveness of combining local feature extraction with the modeling of temporal dependencies.

The trading experiments focus on price response, load curtailment, energy storage scheduling, and parameter sensitivity. To isolate the contributions of individual components, the study employs several benchmark strategies—namely, a no-demand-response scenario, a bid-ranking demand-response mechanism, a deterministic joint optimization approach, and a risk-aware joint optimization approach—with all strategies evaluated under consistent market data and information conditions. Additionally, the energy storage experiments include a scenario without energy storage.

C. Experimental Results

Figure 6 compares the load forecasting results from an ANN, a CNN, an LSTM, and the CNN–LSTM model proposed in this paper for January 1–7, 2020; the panel on the right provides a detailed view of the variations on January 3. All four model curves broadly reproduce the intraday load cycle, but the deviations differ during rapid ramps, peak-demand periods, and changes in the direction of load. The ANN forecast displays pronounced local fluctuations and visible errors during some high-demand intervals. Although the CNN and LSTM models can track the primary trends, they still tend to over- or underestimate local peaks. In contrast, the forecast curve generated by the CNN–LSTM model aligns more closely with the actual observed data, demonstrating superior trend-tracking capability. An examination of a typical day reveals that the model follows the early-morning low-load interval, the subsequent rise in load, and the decline following the evening peak—a result that highlights the effectiveness of combining local feature extraction (via CNN) with the modeling of temporal dependencies (via LSTM). Within the context of the joint electricity procurement and sales optimization framework presented in this paper, more accurate 24-hour-ahead demand forecasting provides a potentially more informative basis for day-ahead electricity procurement declarations and the allocation of demand-response resources, while also helping to minimize the need for real-time adjustments caused by demand-forecasting errors.

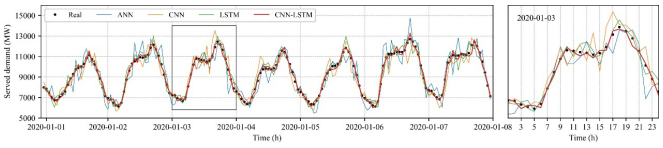


Figure 6: Hourly service load prediction results from different models and a localized comparison for a typical day

Figure 7 illustrates the relationships among electricity procurement, energy storage charge/discharge, state of charge (SoC), and wholesale electricity prices over a 24-hour period. Electricity procurement is generally higher during low-price intervals; specifically, a concentrated charging phase occurs between hours 7 and 10, causing the energy storage SoC to rise and remain at a high level from approximately hour 11 to hour 17. When wholesale electricity prices reach their peak around hours 17–18, the energy storage system switches to discharge mode. Over the following hours, the storage device continues discharging, while external electricity procurement drops significantly, reaching negative values in some intervals. If negative procurement represents net electricity sales in the reported scenario, this indicates that surplus energy is being exported to the grid; otherwise, the definition of this variable should be verified. After electricity prices fall back during hours 23–24, procurement resumes, accompanied by a small amount of charging—demonstrating how energy storage leverages inter-temporal price differentials to optimize the timing of electricity procurement. Within the context of the prediction-driven trading framework presented in this paper, load forecasts establish the demand baseline for procurement planning. Energy storage acts as a flexible resource, coordinating energy allocation across different time slots—subject to energy balance and operational constraints—and, in conjunction with demand-side response, helps manage real-time deviations.

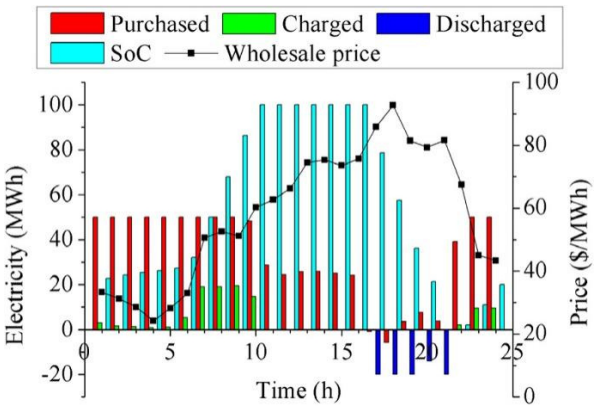


Figure 7: Results of varying wholesale electricity prices

Figure 8 illustrates the relationship between customary electricity consumption, consumption following a demand response, and price signals under various pricing scenarios. Panels (a) through (c) represent three distinct demand response scenarios, while panel (d) displays the aggregate changes in electricity consumption alongside market prices. Overall, electricity consumption after the response tends to shift toward periods with lower prices, while decreasing during periods with higher prices. In scenario (a), consumption is concentrated primarily in the early morning hours, with near-zero usage during the high-price intervals from midday through the evening—demonstrating a significant capacity for shifting load across different time periods. Scenarios (b) and (c) also show an increase in early-morning consumption but

ARCHIVES DES SCIENCES
A Multidisciplinary Journal

retain some demand during the day and evening, reflecting variations in the extent of load adjustment across different scenarios. The aggregate results in panel (d) indicate that, following the response, electricity consumption is notably higher than customary levels during approximately the 2nd to 6th hours, but significantly lower during the afternoon and evening—periods characterized by higher market prices—thereby demonstrating that price incentives can effectively alter the temporal distribution of user electricity consumption.

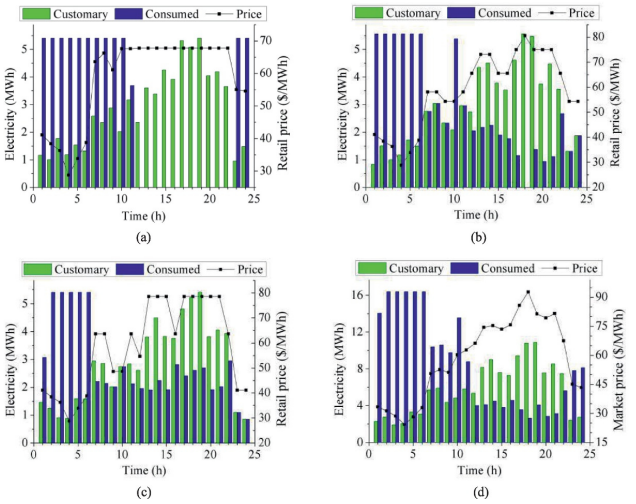


Figure 8: User demand response and aggregated electricity consumption distribution under different pricing scenarios

Figure 9 compares baseline electricity consumption with consumption following demand response across various price scenarios. Panels (a) through (c) illustrate three distinct user-side response scenarios, while panel (d) depicts the relationship between aggregated electricity consumption and market prices.

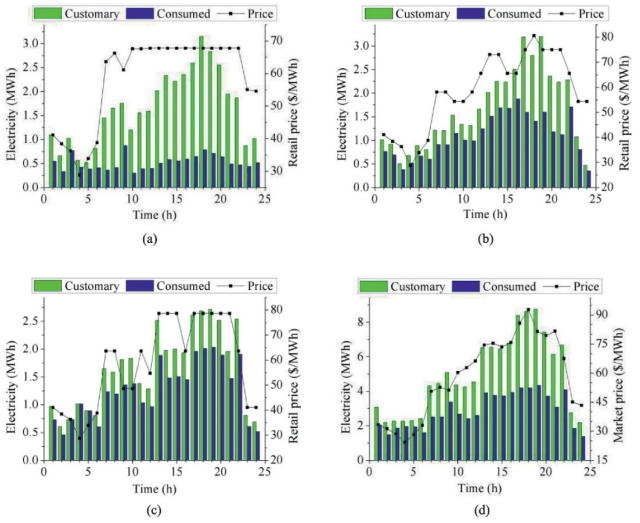


Figure 9: User load-shedding response and changes in aggregated electricity consumption under different price scenarios

Unlike the patterns observed in previous figures—where consumption was notably concentrated during periods of low prices—the results here primarily reflect demand reduction, with no significant evidence of compensatory load shifting. In scenario (a), post-response consumption remains consistently below baseline levels for most of the day, with particularly substantial reductions during the afternoon and early evening, when baseline demand is typically high. In scenario (b), the post-response load profile retains its original daily pattern, though consumption during periods of high demand is noticeably lower. Scenario (c) involves reducing certain peak loads while maintaining a higher level of daytime demand, highlighting the variations in the extent of demand adjustment across different scenarios. The aggregated results in panel (d) show that baseline consumption reaches approximately 8–9 MWh between the 17th and 18th hours, whereas the corresponding post-response consumption is roughly 4 MWh, indicating that the demand response in this scenario effectively lowers the original peak consumption.

Figure 10 illustrates the variations in retail prices, user demand, demand bids, system load, and market settlement results for α values of 0, 15%, and 30%.

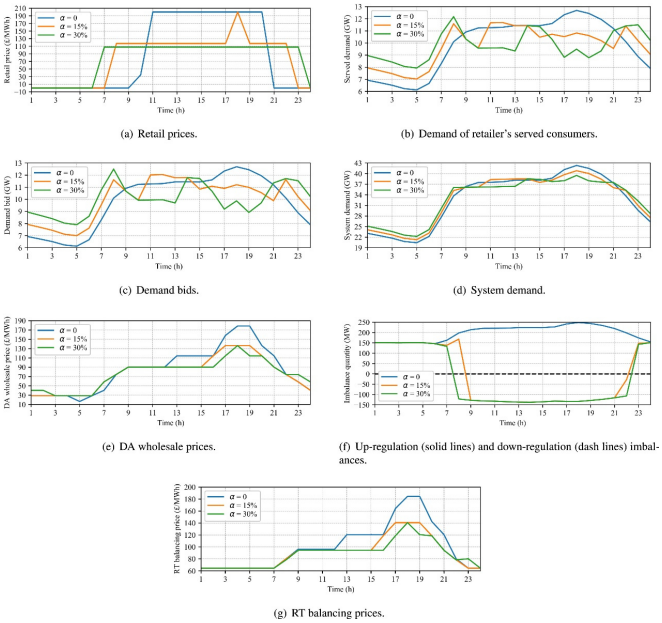


Figure 10: Comparison of real-time market operation results

As shown in Figure 10(a), when $\alpha = 0$, retail prices remain high for an extended period during the day; as α increases, prices generally decrease across most time intervals, though a brief price spike persists when $\alpha = 15\%$, indicating that the impact of this parameter on pricing varies by time period. Panels (b) and (c) demonstrate that user demand and bid curves adjust accordingly: the original evening peak is reduced, with some demand shifting to the morning or other time slots. When $\alpha = 30\%$, the curves exhibit more pronounced fluctuations, suggesting that a larger parameter value does not necessarily lead to a smoother electricity consumption

profile. At the system level, panel (d) shows a decline in the evening demand peak, accompanied by a slight increase in load during the early morning hours, reflecting a temporal reallocation of demand. Correspondingly, the evening peaks in both day-ahead wholesale prices and real-time balancing prices—shown in panels (e) and (g), are reduced, highlighting the interplay between demand adjustments and market price dynamics. Notably, panel (f) reveals that while $\alpha = 0$ is characterized primarily by positive imbalances, $\alpha = 15\%$ and 30% result in significant negative imbalances during the day; these observations show that reductions in peak demand and electricity prices do not necessarily coincide with lower trading imbalances.

Figure 11 illustrates the frequency distributions and pairwise correlations of power samples for time periods 1, 2, and 24. The upper-row histograms display broadly similar right-skewed distributions, with samples primarily concentrated in the 500–2,000 MW range and extending toward higher power levels; this indicates that while each time period shares comparable marginal distribution characteristics, there remains a "tail" of high-power samples. The scatter plots in the bottom row further reveal significant variations in the degree of dependence between different time periods: the correlation coefficient between periods 1 and 2 is 0.9965, with data points clustered tightly around an upward-sloping line, reflecting strong continuity between consecutive periods. In contrast, the correlation coefficients for the pairs (2, 24) and (1, 24) are 0.5671 and 0.5552, respectively; the more dispersed scatter patterns suggest that although a positive association remains across more widely separated periods, the strength of this association is notably weaker. In the context of the load forecasting and risk-aware trading framework proposed in this paper, these findings imply that when constructing demand scenarios and uncertainty sets, it is essential to preserve both the marginal distributions for individual time periods and the inter-temporal dependencies. Relying on independent, hour-by-hour sampling would disrupt the continuity of demand trajectories, potentially compromising day-ahead power procurement, demand response arrangements, and real-time deviation assessments.

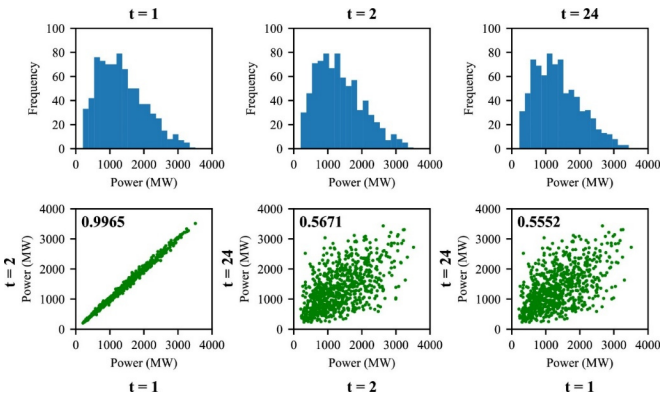


Figure 11: Marginal distribution and cross-period correlation of power samples in typical periods

IV. Conclusion

This study develops a forecast-driven optimization framework for electricity retailers, accounting for demand forecasting errors, uncertainties in user response, and risks associated with the settlement of imbalances. The framework integrates CNN–LSTM-based load forecasting, user demand response, and a two-stage risk-aware decision-making process, with an additional operational scenario incorporating energy storage. Preliminary results indicate that the CNN–LSTM model effectively tracks load fluctuations, while price incentives and energy storage dispatch can adjust the temporal distribution of both demand and electricity procurement. However, peak shaving and a decline in market prices do not necessarily reduce settlement deviations; therefore, the deployment of flexibility resources must holistically consider response costs, delivery reliability, and the gains or losses arising from two-way settlement. The existence of inter-temporal power correlations further underscores the necessity of preserving temporal dependencies when constructing scenarios. As current evidence primarily reflects operational behavior, it is not yet sufficient to quantify overall improvements in profitability and risk mitigation. Future work will utilize user-level response records and consolidated market data to validate the frameworks economic performance, robustness, and operational boundaries through out-of-sample testing and ablation studies.

Author Contributions

Both authors contributed equally to the conception, development, analysis, and preparation of this manuscript. Both authors reviewed the manuscript, approved its final version, and agreed to its submission for publication.

Conflict of Interest

The authors declare that they have no financial or non-financial conflicts of interest in relation to the research presented in this manuscript.

Funding

The authors report that this research received no specific funding from public, commercial, or not-for-profit funding agencies.

Data Availability

No new data were created or analyzed in this study.

References

[1] Sun, B., Wang, F., Xie, J., & Sun, X. (2021). Electricity retailer trading portfolio optimization considering risk assessment in Chinese electricity market. *Electric Power Systems Research, 190*, Article 106833. https://doi.org/10.1016/j.epsr.2020.106833

[2] Qiu, D., Dong, Z., Ruan, G., Zhong, H., Strbac, G., & Kang, C. (2022). Strategic retail pricing and demand bidding of retailers in electricity market: A data-driven chance-constrained programming. *Advances in Applied Energy, 7*, Article 100100. https://doi.org/10.1016/j.adapen.2022.100100

[3] Che, P., Zhang, C., Liu, Y., & Zhang, Y. (2025). An integrated learning and optimization approach to optimal dynamic retail electricity pricing of residential and industrial consumers. *Applied Energy, 382*, Article 125234. https://doi.org/10.1016/j.apenergy.2024.125234

[4] Fabre, A. (2025). *Optimising battery energy storage system trading via energy market operator price forecast* [Preprint]. arXiv. https://arxiv.org/abs/2510.03657

[5] Mohammad, N., & Mishra, Y. (2019). Retailer's risk-aware trading framework with demand response aggregators in short-term electricity markets. *IET Generation, Transmission & Distribution, 13*(13), 2611–2618. https://doi.org/10.1049/iet-gtd.2018.6294

[6] Nojavan, S., Nourollahi, R., Pashaei-Didani, H., & Zare, K. (2019). Uncertainty-based electricity procurement by retailer using robust optimization approach in the presence of demand response exchange. *International Journal of Electrical Power & Energy Systems, 105*, 237–248. https://doi.org/10.1016/j.ijepes.2018.08.041

[7] Teng, J.-H., & Hsieh, C.-H. (2021). Modeling and investigation of demand response uncertainty on reliability assessment. *Energies, 14*(4), Article 1104. https://doi.org/10.3390/en14041104

[8] Huang, C., Li, K., & Zhang, N. (2025). Strategic joint bidding and pricing of load aggregators in day-ahead demand response market. *Applied Energy, 377*, Article 124552. https://doi.org/10.1016/j.apenergy.2024.124552

[9] Jia, X., Ji, P., Chen, F., Deng, H., Chen, M., Zhuang, X., Qi, X., & Tan, Z. (2026). A distributionally robust bilevel optimization model for wholesale–retail electricity market design. *Scientific Reports, 16*, Article 1768. https://doi.org/10.1038/s41598-025-29971-7

[10] Patnam, B. S. K., & Pindoriya, N. M. (2021). Demand response in consumer-centric electricity market: Mathematical models and optimization problems. *Electric Power Systems Research, 193*, Article 106923. https://doi.org/10.1016/j.epsr.2020.106923

[11] Yang, S., Tan, Z., Liu, Z., Lin, H., Ju, L., Zhou, F., & Li, J. (2020). A multi-objective stochastic optimization model for electricity retailers with energy storage system considering uncertainty and demand response. *Journal of Cleaner Production, 277*, Article 124017. https://doi.org/10.1016/j.jclepro.2020.124017

[12] Lu, J., Zhang, Q., Yang, Z., & Tu, M. (2019). A hybrid model based on convolutional neural network and long short-term memory for short-term load forecasting. In *2019 IEEE Power & Energy Society General Meeting (PESGM)* (pp. 1–5). IEEE. https://doi.org/10.1109/PESGM40551.2019.8973549

[13] Dadashi, M., Haghifam, S., Zare, K., Haghifam, M.-R., & Abapour, M. (2020). Short-term scheduling of electricity retailers in the presence of demand response aggregators: A two-stage stochastic bi-level programming approach. *Energy, 205*, Article 117926. https://doi.org/10.1016/j.energy.2020.117926

[14] Jiang, L., Wang, X., Li, W., Wang, L., Yin, X., & Jia, L. (2021). Hybrid multitask multi-information fusion deep learning for household short-term load forecasting. *IEEE Transactions on Smart Grid, 12*(6), 5362–5372. https://doi.org/10.1109/TSG.2021.3091469

[15] Sun, W., Jiang, N., Wang, T., Ma, S., Jin, Y., & Alam, F. (2025). Data-driven distributionally robust collaborative optimization operation strategy for multi-integrated energy systems considers energy trading. *Sustainability, 17*(24), Article 11377. https://doi.org/10.3390/su172411377

[16] Lei, M., Zhang, M., & Wang, K. (2026). Research on bidding optimization strategy for virtual power plants with wind–solar–storage systems based on IGDT–DRO. *Electrical Engineering, 108*(3), Article 208. https://doi.org/10.1007/s00202-026-03544-x