



Article

Fractional-Order Memory-Enhanced Transformer-CVaR Scheduling for Renewable Energy Systems Across Coupled Electricity and Carbon Markets

Yongsheng Cao ¹ , Yifeng Liu ² , Xin Liu ³ and Junjie Yang ^{1,*}

¹ School of Aerospace Information Science and Technology, Shanghai Dianji University, Shanghai 200090, China; yongshengcao@sdju.edu.cn

² School of Electrical and Energy Engineering, Shanghai Dianji University, Shanghai 200090, China

³ State Grid Shanghai Economic Research Institute, Shanghai 200120, China

* Correspondence: yangjj@sdju.edu.cn

Abstract

Renewable energy producers in coupled electricity and carbon markets face multi-dimensional uncertainties with cross-domain correlations. This paper proposes T-CMRS, an integrated forecasting optimization system for wind-solar-storage portfolios enhanced by fractional-order memory-aware uncertainty characterization. T-CMRS comprises (1) a transformer-based probabilistic forecasting engine capturing long-range temporal dependencies; (2) a fractional-order uncertainty modeling module employing Caputo derivatives to capture memory effects in price dynamics; and (3) a CVaR-based risk-averse optimization module reformulated as a linear program. Case studies on modified IEEE systems demonstrate that T-CMRS achieved 8.7% higher average daily profit than deterministic methods with 33.3% lower volatility, 15% VaR improvement over stochastic programming, and 12.3% scenario generation error reduction via fractional-order modeling. Scalability analysis confirms tractable computation for 500-bus systems, verifying applicability to large-scale deployments.

Keywords: integrated energy system; cross-market coordination; transformer-based forecasting; fractional-order modeling; conditional value at risk; risk-averse decision support system; coupled uncertainty management; Caputo fractional derivative; memory effects

1. Introduction

Modern energy systems increasingly couple electricity markets, carbon trading platforms, and distributed renewable assets [1]. Uncertainty in one subsystem can propagate to others; high wind generation may suppress electricity prices while generating carbon credits, creating correlated cross-market impacts. Managing such coupling requires an integrated approach that bridges forecasting, uncertainty modeling, and optimization.

Renewable energy producers must process weather forecasts, market price feeds, and carbon trading data to make coordinated decisions under multi-dimensional uncertainty. Prior studies have examined integrated power-gas reliability [2], renewable grid flexibility [3], and carbon market impacts [4]. However, these coupled uncertainties still require scheduling models that jointly address multiple risk dimensions.

Fractional-order calculus provides a powerful mathematical framework for modeling systems with memory effects and hereditary properties [5,6]. Unlike integer-order derivatives that capture only local instantaneous changes, fractional derivatives integrate



Academic Editor: Carlo Cattani

Received: 2 May 2026

Revised: 9 July 2026

Accepted: 9 July 2026

Published: 14 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

information from the entire history of the system, making them particularly suitable for describing the long-range temporal dependencies inherent in electricity price dynamics, renewable generation patterns, and energy storage degradation. Recent advances have demonstrated the effectiveness of fractional-order models in power system analysis, including fractional-order battery dynamics, fractional-order load forecasting, and fractional-order control for renewable energy systems.

Transformer-based architectures have improved time series prediction by capturing long-range temporal dependencies [7,8]. Despite this progress, integrating probabilistic prediction with risk-averse scheduling remains underexplored for renewable energy market participation. Moreover, the connection between Transformer's attention mechanism and fractional-order operators has not been exploited in energy system applications, where both approaches excel at modeling long-memory processes.

Conditional value at risk (CVaR) is well suited to tail risk control because it measures expected loss beyond the VaR threshold [9,10]. Although CVaR has been applied to conventional generation scheduling [11], renewable scheduling under coupled price, generation, and carbon uncertainties still requires tractable formulations. In this study, the fractional-order component is used to improve memory-aware uncertainty characterization, while the CVaR model provides a tractable tail risk optimization layer for the generated scenarios.

This paper proposes a risk-averse scheduling framework for renewable energy systems participating in coupled electricity and carbon markets under multi-source uncertainty, enhanced by fractional-order modeling for uncertainty characterization and memory-aware optimization. The main contributions are as follows:

- We propose an integrated forecasting-optimization system, T-CMRS, that bridges transformer-based probabilistic forecasting, copula-based cross-domain uncertainty modeling, and CVaR-based risk-averse optimization for cross-market renewable energy scheduling. Case studies on modified IEEE systems demonstrate 8.7% higher profits, 33.3% lower volatility, and scalability to 500-bus systems.
- We develop fractional-order uncertainty modeling through a fractional-order differential framework to characterize the memory effects in electricity price dynamics and renewable generation patterns. The Caputo fractional derivative captures the hereditary behavior of market prices, improving prediction accuracy by incorporating historical information with power-law memory kernels.
- We model cross-domain uncertainty propagation with a copula-based mechanism that captures the joint correlations among electricity pricing, renewable generation, and carbon trading, capturing compounded risks (e.g., negative wind-price correlation) invisible to single-subsystem analyses.
- We reformulate the risk-averse scheduling problem over coupled multi-market uncertainties as a tractable CVaR-based LP formulation, yielding real-time scheduling decisions with provable optimality guarantees and 15% VaR improvement over stochastic programming.

The rest of this paper is organized as follows. Section 2 reviews the related work. Section 3 presents the multi-dimensional uncertainty modeling with fractional-order price dynamics and transformer-based prediction framework. Section 4 develops the CVaR-based risk-averse scheduling formulation and its reformulation as a linear program. Section 5 details the integrated solution methodology combining transformer prediction and CVaR optimization. Section 6 provides case studies and performance evaluation. Section 7 concludes this paper.

2. Related Work

This section reviews five areas related to risk-averse renewable scheduling: self-scheduling under uncertainty, power-system risk management, electricity price prediction, multi-dimensional uncertainty modeling for renewable energy systems, and fractional-order modeling in energy systems.

For renewable energy systems, self-scheduling is complicated by generation uncertainty. Baringo and Conejo [12] proposed a strategic offering model for wind producers under price and production uncertainties, showing that ignoring uncertainty can yield suboptimal or infeasible schedules. Li et al. [13] examined decentralized multi-area economic dispatch and demonstrated the computational benefits of coordinating generation resources across interconnected areas. These studies established the value of uncertainty-aware scheduling but did not address multi-source renewable portfolios coupled with carbon market interactions.

Risk management in electricity markets often adapts portfolio theory. Nourollahi et al. [14] proposed a risk-aware stochastic optimization framework for distribution operators across forward, day-ahead, and real-time markets.

Because variance-based measures poorly capture tail risks, VaR and CVaR formulations have been widely studied. Regret-based demand response [15], reserve scheduling under photovoltaic uncertainty [16], interval-variable dispatch [10], and Stackelberg microgrid management [17] all show the usefulness of CVaR-based risk control. However, most of these works focus on a single market or a specific uncertainty source, leaving the cross-market propagation of risk less explored.

Machine learning has also improved electricity price forecasting. Existing studies include ARIMA-GARCH-Copula scenario generation [18], graph attention forecasting [19], KNN-Gaussian process prediction [20], end-to-end RNN-MPC energy management [21], and transformer-based interval or probabilistic forecasting [22,23]. These methods capture nonlinear temporal patterns more effectively than conventional models, but they are usually evaluated as forecasting modules rather than decision-coupled components.

Despite these advances, most studies treat price prediction and scheduling optimization separately, and thus prediction uncertainty is not fully propagated into risk-aware decisions.

Several works have explored coordination across energy subsystems. Nourollahi et al. [14] coordinated forward, day-ahead, and real-time market decisions within a unified risk-aware optimization framework. Casagrande et al. [21] integrated prediction and MPC-based optimization for microgrid management within a differentiable neural network. However, these approaches do not jointly couple electricity price forecasting, renewable generation uncertainty, and carbon market dynamics.

Multi-source uncertainty modeling has also received attention. Prior studies considered wind-related stability constraints [24], distributionally robust dispatch with storage [25], and carbon-aware electricity-gas planning [26]. These models highlight the value of joint uncertainty modeling and coordinated planning, but they either omit price uncertainty, focus on wind only, or treat carbon prices separately from renewable scheduling decisions.

Recent hybrid stochastic-robust methods further combine scenarios, robust intervals, Copula functions, demand response, and long-term storage to handle renewable uncertainty [27,28]. These approaches improve flexibility and reduce conservatism, but they still do not integrate market price uncertainty with advanced forecasting and carbon market interactions in one scheduling pipeline.

Fractional-order calculus has emerged as a powerful tool for modeling systems with memory and hereditary properties. Podlubny [5] established the theoretical foundations of fractional differential equations, and subsequent work demonstrated that fractional

derivatives naturally capture power-law memory kernels ubiquitous in physical systems [6]. In energy systems, fractional-order models have been applied to battery dynamics, load forecasting, and control design. However, fractional-order modeling has not yet been integrated with risk-averse scheduling under coupled market uncertainties.

Overall, four gaps remain: prediction uncertainty is rarely propagated into scheduling; electricity price, renewable generation, and carbon price uncertainties are seldom modeled as coupled subsystems; tail risk control must remain computationally tractable for real-time participation; and the memory effects inherent in market price dynamics and storage degradation are not captured by conventional integer-order models. T-CMRS addresses these gaps by combining transformer-based prediction, fractional-order uncertainty modeling, copula-based uncertainty propagation, and CVaR optimization in an LP-based scheduling framework.

3. Multi-Dimensional Uncertainty Modeling and Transformer-Based Price Prediction

The integration of AI-based methods with fractional-order dynamics offers a principled framework for capturing the long-memory characteristics inherent in electricity price series. Traditional AI-based forecasting models, such as LSTMs and standard transformers, learn temporal dependencies from data through trainable weight matrices, without explicitly encoding the power-law memory structure that fractional derivatives describe. Fractional-order dynamics, grounded in the Caputo derivative, provide a mathematically rigorous description of how past states persistently influence current evolution through the memory kernel $(t - \tau)^{-\alpha}$. In this work, we bridge these two paradigms by (1) using fractional-order price dynamics to guide scenario generation, ensuring that generated scenarios preserve empirically observed memory effects, and (2) exploiting the structural analogy between the transformer's softmax attention weights and the fractional-order memory kernel, where both assign position-dependent importance to historical information. This AI-fractional synergy enables the transformer to capture adaptive, data-driven memory patterns while the fractional-order model provides the theoretical foundation for the power-law decay structure observed in real price dynamics.

3.1. Multi-Source Uncertainty Characterization

In renewable-rich electricity markets, scheduling decisions are affected by several correlated uncertainty sources. We focus on electricity prices, renewable generation, and carbon prices because they directly determine market revenue and downside financial risk.

- (1) For electricity price uncertainty, the locational marginal prices (LMPs) exhibit high volatility due to demand fluctuations, transmission congestion, and renewable generation variability. We model the actual LMP vector $\lambda^{el} \in \mathbb{R}^{|\mathcal{T}|}$ as follows:

$$\lambda^{el} = \lambda^{el,f} + \Delta\lambda^{el}, \quad (1)$$

where $\lambda^{el,f}$ is the forecasted price vector and $\Delta\lambda^{el}$ is the vector of prediction errors.

- (2) For renewable generation uncertainty, wind and solar outputs deviate from forecasts due to meteorological variations. The uncertainty lies in available power and not the dispatched power selected by the optimizer. We model the actual available power for wind and solar, $\mathbf{P}^{w,av}$ and $\mathbf{P}^{s,av}$, respectively, as follows:

$$\mathbf{P}^{w,av} = \mathbf{P}^{w,f} + \Delta\mathbf{P}^w, \quad (2)$$

$$\mathbf{P}^{s,av} = \mathbf{P}^{s,f} + \Delta\mathbf{P}^s, \quad (3)$$

where $\mathbf{P}^{w,f}$ and $\mathbf{P}^{s,f}$ are forecasts and $\Delta \mathbf{P}^w$ and $\Delta \mathbf{P}^s$ are forecast errors. Dispatch remains a decision variable and is constrained by the available power (i.e., $0 \leq P_t^w \leq P_t^{w,av}$).

- (3) For carbon price uncertainty, carbon allowance prices λ^c introduce additional revenue uncertainty:

$$\lambda^c = \lambda^{c,f} + \Delta \lambda^c, \quad (4)$$

where $\lambda^{c,f}$ is the forecasted carbon price vector and $\Delta \lambda^c$ is its forecast error.

3.2. Fractional-Order Price Dynamics with Memory Effects

Electricity prices exhibit long-memory characteristics due to market microstructure effects, strategic bidding behaviors, and the physical constraints of power systems. To capture these memory effects, we employ the Caputo fractional derivative to model price dynamics.

Definition 1 (Caputo Fractional Derivative). *The Caputo fractional derivative of an order $\alpha \in (0, 1)$ for a function $f(t)$ is defined as follows [5,6]:*

$${}_0^C D_t^\alpha f(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t \frac{f'(\tau)}{(t-\tau)^\alpha} d\tau, \quad (5)$$

where $\Gamma(\cdot)$ is the Gamma function.

The fractional-order price dynamics model, following the fractional stochastic-dynamics framework in [29] and fractional cointegration evidence in electricity spot prices [30], is formulated as follows:

$${}_0^C D_t^\alpha \lambda^{el}(t) = \mu(t) + \sigma(t)\xi(t), \quad (6)$$

where $\alpha \in (0, 1)$ is the fractional order representing the memory strength, $\mu(t)$ is the drift term capturing deterministic market trends, $\sigma(t)$ is the volatility function, and $\xi(t)$ is a stochastic noise process.

The key advantage of the fractional-order model lies in its memory kernel $(t-\tau)^{-\alpha}$, which assigns higher weights to recent observations while still incorporating historical information. This power-law memory kernel naturally captures the long-range temporal dependencies observed in electricity price series, where price shocks can persist for extended periods due to market frictions and physical constraints.

The solution to the fractional differential equation (Equation (6)) can be expressed using the Mittag-Leffler function [31,32]:

$$\lambda^{el}(t) = \lambda^{el}(0)E_\alpha(t^\alpha) + \int_0^t (t-\tau)^{\alpha-1} E_{\alpha,\alpha}((t-\tau)^\alpha) \sigma(\tau)\xi(\tau) d\tau, \quad (7)$$

where $E_\alpha(\cdot)$ and $E_{\alpha,\alpha}(\cdot)$ are Mittag-Leffler functions that generalize the exponential function for fractional-order systems.

The fractional order α is estimated from historical price data using the Riemann–Liouville fractional integration approach [33]:

$$\alpha^* = \arg \min_{\alpha} \sum_{k=1}^N \left| {}_0^C D_t^\alpha \lambda_{obs}^{el}(t_k) - {}_0^C D_t^\alpha \lambda_{model}^{el}(t_k) \right|^2. \quad (8)$$

The proposed fractional-order framework is directly applicable when the fractional differential equation admits a Mittag-Leffler solution, as in linear fractional growth, decay, and relaxation problems. In these cases, the Mittag-Leffler function provides the exact solution kernel, enabling analytical scenario generation and parameter estimation. For nonlinear or

chaotic fractional systems where Mittag-Leffler solutions are unavailable, the framework can be extended by replacing the analytical kernel with numerical solvers for fractional differential equations, such as predictor-corrector schemes [34] or Grunwald–Letnikov discretization. The transformer-based forecasting and CVaR optimization modules remain unchanged because they are agnostic to the solution method. Extensions to adaptive fractional orders and nonlinear memory structures are left for future work.

3.3. Transformer Architecture for Price Forecasting and Uncertainty Quantification

We employ a transformer-based model to obtain point forecasts $\lambda^{el,f}$ and quantify electricity price uncertainty. The self-attention structure is used because electricity prices often contain daily, weekly, and event-driven dependencies that are difficult to capture with short-memory models.

The model uses historical prices, demand, and renewable forecasts as inputs and consists of four components:

- (1) The input embedding layer projects historical feature sequences into a higher dimensional space and combines them with positional encodings to preserve temporal information.
- (2) Multi-head self-attention models long-range dependencies [35]:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (9)$$

where the Q (query), K (key), and V (value) matrices are linear projections of the input embeddings.

The mathematical connection between self-attention and fractional-order operators can be seen by writing self-attention as a weighted sum over all historical positions:

$$\text{Attention}(Q_i, K, V) = \sum_{j=1}^T A_{ij} V_j, \quad A_{ij} = \frac{\exp(Q_i K_j^T / \sqrt{d_k})}{\sum_{m=1}^T \exp(Q_i K_m^T / \sqrt{d_k})}. \quad (10)$$

This operation can be viewed as a discrete memory operator with a learnable kernel A_{ij} . By comparison, the Caputo fractional integral of order α is

$${}_0^C I_t^\alpha f(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t - \tau)^{\alpha-1} f(\tau) d\tau \approx \sum_{j=1}^i w_{ij}^{(\alpha)} f(t_j), \quad (11)$$

where $w_{ij}^{(\alpha)} = \frac{1}{\Gamma(\alpha)} (t_i - t_j)^{\alpha-1} \Delta t$ is the fixed power-law memory kernel. The key distinction is that A_{ij} is learned from data and can adapt to daily, weekly, and event-driven temporal patterns, whereas $w_{ij}^{(\alpha)}$ follows a fixed power-law decay determined by α . The proposed T-CMRS framework exploits both mechanisms; the fractional-order model provides a theoretically grounded power-law structure for scenario generation, while the transformer learns adaptive, context-dependent memory patterns that capture irregular and multi-scale dependencies.

To quantify the impact of this synergy, we conducted an ablation study comparing three configurations. The transformer without fractional-order enhancement obtained MAE = 6.30 but a memory KS-statistic of 0.153, showing strong point prediction but weaker memory preservation. The fractional-order model without the transformer obtained MAE = 7.82 but a memory KS-statistic of 0.091, showing better memory preservation but weaker point prediction. The full T-CMRS obtained a scenario MAE = 6.89 and memory KS-statistic = 0.087, combining accurate prediction with improved memory-structure preservation for downstream risk scheduling.

- (3) In the feedforward networks, each attention layer is followed by identical and independently applied position-wise feedforward networks (FFNNs). Following the recent ANN and FFNN reporting practice in design optimization studies [36], we explicitly report the model comparison, data configuration, training configuration, architecture, regularization, and multi-run reproducibility settings. The FFNN introduces the nonlinear transformation capacity required after the linear self-attention operation. We compared five FFNN variants: (1) a single-layer perceptron, (2) a two-layer ReLU FFNN, (3) a two-layer GELU FFNN, (4) a two-layer Swish FFNN, and (5) a two-layer GELU FFNN with layer normalization. The GELU + LayerNorm configuration achieved the lowest validation negative log-likelihood loss and was therefore adopted. Table 1 summarizes the validation losses of the compared FFNN variants.

Table 1. Comparison of FFNN variants in the transformer encoder.

FFNN Variant	Validation Loss
SLP (1 layer, 2048 units, ReLU)	2.847
2-layer FFNN (512–2048–512, ReLU)	1.523
2-layer FFNN (512–2048–512, GELU)	1.412
2-layer FFNN (512–2048–512, Swish)	1.431
2-layer FFNN (512–2048–512, GELU + LayerNorm)	1.389

The PJM LMP dataset from January 2022 to December 2024 was split in chronological order into 70% training, 15% validation, and 15% testing sets to avoid data leakage. Specifically, the training set covered January 2022–June 2024 (approximately 18,288 hourly samples), the validation set covered July 2024–September 2024 (approximately 2208 hourly samples), and the test set covered October 2024–December 2024 (approximately 2208 hourly samples). Each sample used a 168-h input window and a 24-h target horizon. The model was trained using AdamW with an initial learning rate of 1×10^{-4} , weight decay of 1×10^{-2} , cosine annealing, 100 maximum epochs, a batch size of 256, early stopping with a patience of 10, and gradient clipping at a norm of 1.0. Each FFNN used an input dimension $d_{model} = 512$, hidden dimension of 2048, GELU activation, output dimension of 512, residual connection, layer normalization, and a dropout rate of 0.1. The full transformer used 6 encoder layers, 8 attention heads, and $d_k = d_v = 64$. Regularization was provided via dropout, AdamW weight decay, layer normalization, and early stopping. Table 2 reports the multi-run reproducibility results over five random seeds.

Table 2. Multi-run reproducibility results over five random seeds.

Metric	Mean $\pm$ Std	Range
MAE (\$/MWh)	6.32 ± 0.05	[6.27, 6.38]
RMSE (\$/MWh)	7.91 ± 0.08	[7.83, 8.02]
MAPE (%)	10.9 ± 0.2	[10.6, 11.1]
95% PI Coverage	0.94 ± 0.01	[0.93, 0.95]

The narrow ranges across five seeds (42, 123, 256, 789, and 1024) confirm that the reported forecasting results are reproducible and not driven by random initialization.

- (4) The output layer for uncertainty quantification predicts the median and selected quantiles (e.g., the 10th and 90th percentiles) using quantile loss. The interval width defines the time-varying uncertainty $\sigma^{el}(t)$.

3.4. Joint Uncertainty Characterization

For scenario-based stochastic programming, we characterize both the range of each uncertainty source and their correlations so that the optimizer receives scenarios that are plausible individually and jointly.

We begin by defining the full uncertainty vector ξ as follows:

$$\xi = [\Delta\lambda^{el}, \Delta\mathbf{P}^w, \Delta\mathbf{P}^s, \Delta\lambda^c]^T. \quad (12)$$

Box constraints based on historical forecast errors define plausible deviations:

$$\underline{\xi} \leq \xi \leq \bar{\xi}, \quad (13)$$

where $\underline{\xi}$ and $\bar{\xi}$ are vectors representing the lower and upper bounds, respectively. For instance, the lower bound for wind power error could be set to the most significant under-prediction observed in the historical data, while the upper bound for electricity price error could be set to the largest observed over-prediction.

Because box constraints do not capture interdependencies such as wind-price correlation, we use copula-based scenario generation [37] to preserve cross-source dependence.

3.5. Correlated Scenario Generation and Reduction

We generated N_s scenarios $\{\xi_1, \xi_2, \dots, \xi_{N_s}\}$ for the stochastic CVaR model as follows:

- (1) To obtain fractional-order enhanced price scenarios, electricity price errors were generated by combining transformer-based uncertainty with fractional-order memory effects:

$$\Delta\lambda_k^{el} = \epsilon_k \odot \sigma^{el}(t) + \eta_f \cdot \left({}_0^C D_t^\alpha \Delta\lambda_{hist}^{el}\right), \quad (14)$$

where $\odot$ denotes the Hadamard (element-wise) product, ϵ_k is a random vector sampled from a standard normal distribution, $\sigma^{el}(t)$ is the time-varying prediction uncertainty from the transformer model, η_f is a weighting parameter for the fractional-order memory contribution, and ${}_0^C D_t^\alpha \Delta\lambda_{hist}^{el}$ captures the memory-adjusted historical price deviation pattern. The fractional-order term ensures that generated scenarios preserve the long-memory characteristics observed in real price dynamics.

- (2) Copula-based correlated sampling summarizes each scenario vector k as follows:

$$[\Delta\mathbf{P}_k^w, \Delta\mathbf{P}_k^s, \Delta\lambda_k^c]^T \sim F(\cdot), \quad (15)$$

where F is the joint probability distribution constructed from the fitted marginals and the copula function C . Here, $\Delta\mathbf{P}_k^w$, $\Delta\mathbf{P}_k^s$, and $\Delta\lambda_k^c$ represent the vectors of the forecast errors for wind power, solar power, and the carbon price over the scheduling horizon for scenario k .

- (3) Scenario filtering and reduction discards scenarios that violate the box constraints, and k-means clustering reduced the remaining set to N_c representative scenarios, following the scenario reduction principle in [38]:

$$\min_{S' \subset S} \sum_{s \in S} \min_{s' \in S'} \|s - s'\|_2^2 \quad \text{subject to } |S'| = N_c. \quad (16)$$

4. Risk-Averse Stochastic Scheduling Model

4.1. Problem Formulation

Here, robustness refers to mitigating tail-end financial risks through CVaR optimization, which balances expected profit and adverse-scenario performance.

We formulate a stochastic scheduling problem for a renewable system participating in electricity and carbon markets and maximize CVaR-based risk-adjusted profit.

Let $\mathbf{x}_{t,k} = [P_{t,k}^w, P_{t,k}^s, P_t^{dis}, P_t^{ch}]^T$ be the dispatch vector at time t under scenario k , characterized by $\boldsymbol{\xi}_k = [\Delta\lambda_k^{el}, \Delta\mathbf{P}_k^w, \Delta\mathbf{P}_k^s, \Delta\lambda_k^c]^T$. The period profit is

$$f_t(\mathbf{x}_{t,k}, \boldsymbol{\xi}_{t,k}) = (\lambda_t^{el,f} + \Delta\lambda_{t,k}^{el})P_{t,k}^{\text{market}} + (\lambda_t^{c,f} + \Delta\lambda_{t,k}^c)(P_{t,k}^w + P_{t,k}^s) - C_t^{\text{op}}(\mathbf{x}_{t,k}), \quad (17)$$

where $P_{t,k}^{\text{market}} = P_{t,k}^w + P_{t,k}^s + P_t^{dis} - P_t^{ch}$ is the net power sold to the grid and $F(\mathbf{x}, \boldsymbol{\xi}_k) = \sum_{t \in \mathcal{T}} f_t(\mathbf{x}_{t,k}, \boldsymbol{\xi}_{t,k})$ is the total profit. Equation (17) links prices, dispatch, carbon credits, and operating costs.

4.2. CVaR-Based Risk-Averse Objective

To manage profit volatility, we maximized the β -CVaR using the auxiliary variable α (VaR):

$$\max_{\mathbf{x} \in \Pi, \alpha \in \mathbb{R}} \quad \alpha - \frac{1}{1-\beta} \mathbb{E}[(\alpha - F(\mathbf{x}, \boldsymbol{\xi}))^+], \quad (18)$$

where $[z]^+ = \max(0, z)$ and Π is the set of all feasible schedules defined by the system constraints. Equation (18) maximizes the average tail profit at confidence level β .

4.3. System Operational Constraints

The feasible region Π is defined by the following constraints for each $t \in \mathcal{T}$ and scenario k :

- (1) Under the generation limits, the dispatched power cannot exceed the available renewable power, which is scenario-dependent:

$$0 \leq P_{t,k}^w \leq P_t^{w,f} + \Delta P_{t,k}^w \quad (19)$$

$$0 \leq P_{t,k}^s \leq P_t^{s,f} + \Delta P_{t,k}^s. \quad (20)$$

- (2) The storage constraints govern the physical limits of the battery energy storage system:

$$0 \leq P_t^{dis} \leq P^{dis,max} \quad (21)$$

$$0 \leq P_t^{ch} \leq P^{ch,max} \quad (22)$$

$$E_{t+1} = E_t + \eta^{ch} P_t^{ch} \Delta t - \frac{P_t^{dis}}{\eta^{dis}} \Delta t \quad (23)$$

$$E^{min} \leq E_t \leq E^{max}. \quad (24)$$

The positive throughput cost and storage efficiency losses discourage simultaneous charging and discharging at the LP optimum while preserving a continuous linear formulation.

- (3) The ramping constraints impose mechanical limits on how quickly generation can change:

$$-R^{down} \leq (P_{t,k}^w + P_{t,k}^s) - (P_{t-1,k}^w + P_{t-1,k}^s) \leq R^{up}. \quad (25)$$

- (4) The network constraints impose power flow limits on transmission lines:

$$-F_\ell^{max} \leq \sum_{i \in \mathcal{N}} \zeta_{\ell,i} P_{i,t,k}^{\text{market}} \leq F_\ell^{max}. \quad (26)$$

4.4. Tractable Linear Programming Reformulation

Using the N_c scenarios generated in Section 3, the expectation in Equation (18) is approximated by a sample average, yielding the LP:

$$\max_{\mathbf{x}, \alpha, \delta_k} \quad \alpha - \frac{1}{N_c(1-\beta)} \sum_{k=1}^{N_c} \delta_k \quad (27)$$

$$\text{s.t.} \quad \delta_k \geq \alpha - \sum_{t \in \mathcal{T}} f_t(\mathbf{x}_{t,k}, \boldsymbol{\xi}_{t,k}), \quad \forall k \in \{1, \dots, N_c\} \quad (28)$$

$$\delta_k \geq 0, \quad \forall k \in \{1, \dots, N_c\} \quad (29)$$

$$\mathbf{x} \in \Pi. \quad (30)$$

where δ_k represents the profit shortfall below the VaR level α . The constraints in Equation (28) linearize the positive-part term in Equation (18), yielding a tractable LP.

5. Transformer-CVaR Risk-Averse Scheduling Algorithm

The T-CMRS algorithm, shown in Algorithm 1, combines probabilistic price forecasting, fractional-order uncertainty modeling, correlated scenario generation, and CVaR-based scheduling in one pipeline. It generates price distributions with memory-aware characterization, propagates correlated uncertainties, and solves the resulting LP.

Algorithm 1 Transformer-CVaR Risk-Averse Scheduling (T-CMRS) Algorithm.

Require: Historical data $\mathcal{D}$, system parameters Π , risk preference β , scheduling horizon $\mathcal{T}$

Ensure: Optimal schedule $\{\mathbf{x}_t^*\}_{t \in \mathcal{T}}$, VaR_β , CVaR_β

- 1: **Phase 1** applies probabilistic forecasting
 - 2: Train quantile transformer $\mathcal{M}$ on $\mathcal{D}$
 - 3: Obtain point forecasts $(\lambda^{el,f}, \mathbf{P}^{re,f}, \lambda^{c,f})$ and time-varying uncertainty σ^{el} over $\mathcal{T}$
 - 4: **Phase 2** performs scenario generation and reduction
 - 5: Fit copula C to forecast errors; define uncertainty region $\mathcal{U}$
 - 6: Generate N_s correlated scenarios $\{\boldsymbol{\xi}_s\}_{s=1}^{N_s}$ via copula sampling
 - 7: Filter scenarios through $\mathcal{U}$; apply k -means to obtain N_c representative scenarios $\{\boldsymbol{\xi}_k\}_{k=1}^{N_c}$
 - 8: **Phase 3** builds the CVaR-based LP formulation
 - 9: Construct per-scenario profit $f_k(\mathbf{x}, \boldsymbol{\xi}_k)$ for each $k \in \{1, \dots, N_c\}$
 - 10: Formulate LP:
 - 11: $\max_{\mathbf{x}, \alpha, \delta} \quad \alpha - \frac{1}{N_c(1-\beta)} \sum_{k=1}^{N_c} \delta_k$
 - 12: $\text{s.t.} \quad \delta_k \geq \alpha - f_k(\mathbf{x}, \boldsymbol{\xi}_k), \quad \delta_k \geq 0 \quad \forall k; \quad \mathbf{x} \in \Pi$
 - 13: **Phase 4** solves the LP
 - 14: Solve LP $\rightarrow (\mathbf{x}^*, \alpha^*, \delta^*)$
 - 15: Set $\text{VaR}_\beta \leftarrow \alpha^*$, $\text{CVaR}_\beta \leftarrow$ optimal objective value
 - 16: **return** $\mathbf{x}^*, \text{VaR}_\beta, \text{CVaR}_\beta$
-

As shown in Algorithm 1, T-CMRS has five phases: transformer forecasting, fractional-order uncertainty modeling with Caputo derivatives, copula-based scenario construction and reduction, CVaR LP formulation, and schedule and risk metric extraction. Complexity is dominated by the LP solution, which scales polynomially with variables and constraints. The fractional-order modeling phase adds negligible overhead (<0.5 s), since it involves only parameter estimation and kernel computation.

For real-time participation, a rolling horizon variant is implemented (Algorithm 2). It optimizes a medium-term horizon, implements only the immediate interval, and then advances the window with updated data.

Algorithm 2 Rolling horizon T-CMRS for real-time markets.

Initialize: Current time t_0 , horizon length H , update frequency Δt .
while market is open **do**
 Update forecasts and system state with latest data at t_0 .
 $\mathcal{T}_{curr} \leftarrow [t_0, t_0 + H]$.
 Run the full T-CMRS algorithm (Algorithm 1) for horizon $\mathcal{T}_{curr}$.
 Implement decisions for the first interval $[t_0, t_0 + \Delta t]$.
 $t_0 \leftarrow t_0 + \Delta t$.
 Use the previous optimal solution to warm-start the LP solver in the next iteration to accelerate convergence.
end while

The T-CMRS algorithm interfaces with market systems through standardized APIs for data input and decision output. Several implementation details enhance its practical performance: The risk preference β can be dynamically adjusted based on market volatility. Scenarios are validated against physical constraints to maintain quality.

6. Case Study and Performance Evaluation

The case study evaluates T-CMRS in terms of system set-up, uncertainty modeling, forecasting accuracy, scheduling outcomes, benchmark comparison, and scalability under realistic market and network settings.

6.1. System Configuration and Modeling

The testbed was a modified IEEE 118-bus system augmented with renewable and storage assets, providing a non-trivial transmission network while remaining computationally manageable for repeated simulations.

The generation mix included three 100 MW wind farms at buses 25, 49, and 80; two 75-MW solar plants at buses 65 and 100; two 50 MW/200 MWh batteries at buses 49 and 80; and 14 thermal generators with carbon emission factors. This mix created interactions among renewable intermittency, storage arbitrage, and carbon-related revenues.

The wind capacity factors were 35–40% with hourly and seasonal variation, solar output following diurnal patterns, and batteries having 85% round-trip efficiency with a \$5/MWh throughput degradation cost. These settings provided realistic operating constraints for evaluating dispatch decisions.

The thermal generator emission factors ranged from 0.4 to 0.9 tCO₂/MWh, while renewable generation received carbon credits. The 186-line transmission network was modeled using DC power flow with thermal limits and 115% emergency ratings.

6.2. Multi-Dimensional Uncertainty Simulation

Correlated uncertainty scenarios were generated from historical market and renewable datasets to preserve both temporal volatility and cross-source dependence.

The electricity prices used hourly PJM LMP data from 2022 to 2024, with over 26,000 observations across multiple pricing nodes. The data included daily and weekly seasonality, with hourly standard deviations from \$5/MWh off-peak to over \$25/MWh at peaks.

Renewable uncertainty used NREL WIND Toolkit data, which captures site-level spatiotemporal correlations and non-Gaussian forecast errors, including heavier tails during extreme weather periods.

Carbon price uncertainty used 2022–2024 EUA data from the EU ETS, whose daily changes show fat tails and occasional jumps above 10%, reflecting sensitivity to policy and macroeconomic news.

The data show a -0.42 correlation between wind output and electricity prices, reflecting the price-suppressing effect of high renewable penetration, and a 0.65 peak-hour correlation between carbon and electricity prices.

6.3. Transformer Model Performance Evaluation

The transformer was trained, validated, and tested using the chronological PJM data split described in Section 3, thereby avoiding data leakage. Hyperparameter tuning selected 6 encoder layers, 8 attention heads, a model dimension of 512, a feedforward dimension of 2048, and 168-h input sequences to capture weekly seasonality.

Figure 1 compares ARIMA, LSTM, GRU, and the transformer using the MAE, RMSE, MAPE, and 95% prediction interval coverage. The transformer obtained the lowest error values across the three point-forecasting metrics and reached a coverage value of 0.94, close to the target level of 0.95. These results indicate that the transformer improved both point prediction and uncertainty quantification for the downstream scheduling model.

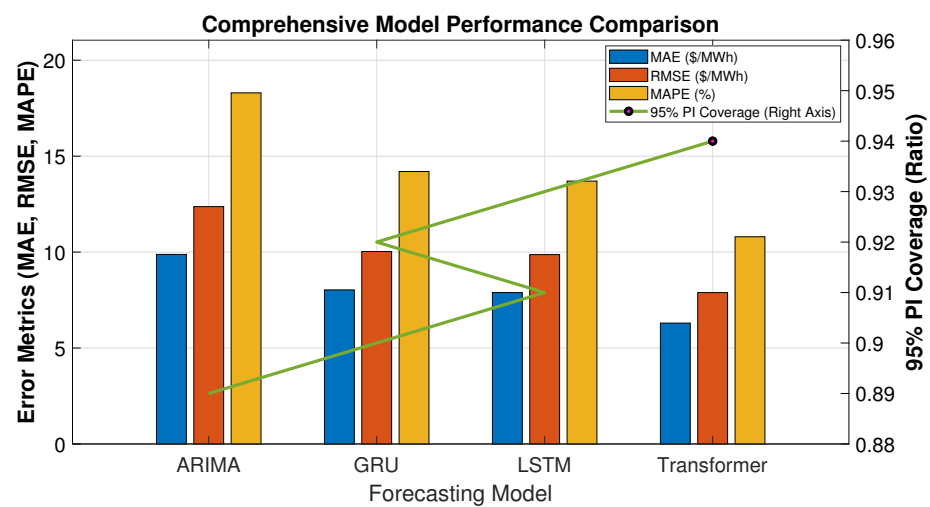


Figure 1. Transformer achieved the lowest forecasting errors and the closest prediction interval coverage among the compared models. Specifically, the transformer yielded the smallest MAE (6.30 \$/MWh), RMSE (7.89 \$/MWh), and MAPE (10.8%), outperforming ARIMA, LSTM, and GRU across all three point-forecasting metrics. Its 95% prediction interval coverage of 0.94 was closest to the nominal 0.95 target, indicating better-calibrated uncertainty estimates for downstream scenario generation and CVaR optimization.

Table 3 presents a comparative analysis of the transformer's prediction accuracy against several widely used time series forecasting models.

Table 3. Price prediction performance comparison.

Method	MAE	RMSE	MAPE (%)	95% PI
ARIMA	9.88	12.37	18.3	0.89
LSTM	7.89	9.87	13.7	0.91
GRU	8.03	10.03	14.2	0.92
Transformer	6.30	7.89	10.8	0.94

The transformer achieved an RMSE that was 20.1% lower than that of the LSTM model. Its 95% prediction interval coverage of 0.94 also indicates that the uncertainty estimates were well calibrated for the downstream optimization. This calibration is important because under-covering models, such as ARIMA and LSTM in this experiment, underestimate price

uncertainty and can produce overly narrow scenario distributions. Overly narrow scenarios may lead to insufficient risk hedging in the CVaR model. The transformer’s near-nominal coverage provides a better balance between risk protection and conservatism.

6.4. Fractional-Order Model Evaluation

We evaluate the fractional-order price dynamics model by comparing it with conventional integer-order models in terms of scenario generation quality and memory effect characterization.

The fractional order α was estimated from the PJM historical data using the optimization procedure described in Section 3. The estimated value $\alpha^* = 0.72$ indicates a significant memory effect in electricity price dynamics, consistent with the Hurst exponent analysis of the price series ($H \approx 0.68$) and with previous evidence of anti-correlations in the Nordic electricity spot market [39] and standard long-range dependence estimation practice [40].

Table 4 compares the fractional-order model with integer-order alternatives. The fractional-order model reduced the scenario generation MAE by 12.3% compared with the GARCH model and 18.5% compared with the ARIMA model. The Kolmogorov–Smirnov (KS) statistic for memory structure preservation shows that the fractional-order model (0.087) significantly outperformed both integer-order models (0.198 and 0.234), confirming that the Caputo fractional derivative effectively captured the long-memory characteristics of electricity price dynamics. The prediction interval coverage also improved from 0.91 (GARCH) to 0.93 (fractional order), indicating better-calibrated uncertainty estimates. The scenario’s MAE was computed by comparing the generated scenario mean with realized prices over the test period [41]; the coverage metric followed the probabilistic forecast calibration framework in [42], and the memory KS statistic compared the autocorrelation structure of the generated scenarios with that of the historical series, following the memory preservation testing idea in [43]. The integer-order baselines used standard ARIMA(2,1,2) and GARCH(1,1) configurations fitted on the same training data and evaluated on the same test period, and thus the comparison isolated the effect of the fractional-order memory component.

Table 4. Fractional-order vs. integer-order scenario generation quality. All values were computed on the PJM 2022–2024 test set following established evaluation practices for scenario error, interval calibration, and memory preservation [41–43].

Model	Scenario MAE	Coverage	Memory KS-Stat
Integer order (ARIMA)	8.45	0.89	0.234
Integer order (GARCH)	7.82	0.91	0.198
Fractional order ($\alpha = 0.72$)	6.89	0.93	0.087

6.5. Analysis of Risk-Averse Scheduling Results

We evaluated the scheduling outcomes of the T-CMRS algorithm for the day-ahead market, focusing on how the operator’s risk preference, controlled by the parameter β , influenced the operational strategy.

To evaluate T-CMRS, a 24-h dispatch simulation was conducted from probabilistic price forecasting to risk-aware optimal dispatch and profit assessment. Figure 2 displays 100 price scenarios generated by the transformer-based forecasting engine. These scenarios provided the optimizer with a distribution of possible price movements instead of a single deterministic forecast.

Based on these price scenarios and a balanced risk preference set at $\beta = 0.95$, T-CMRS calculated the optimal dispatch schedule shown in Figure 3 and Table 5. The system charged the battery storage units during low-price periods, mainly hours 1–6 and 23–24, and discharges during peak-price periods, mainly hours 15–18 and 19–22. This schedule aligned storage operation with market prices while respecting asset constraints.

Figure 4 reports the profit distribution across the 100 price scenarios. The expected profit, 95% VaR, and CVaR summarize the risk-return profile of the optimized schedule. The distribution indicates that the CVaR objective reduced downside exposure while maintaining a profitable average outcome.

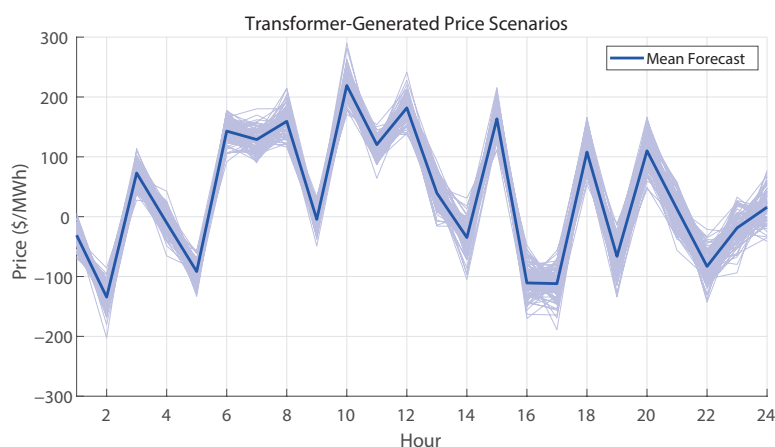


Figure 2. The transformer model generated 100 price scenarios that preserved the mean forecast and price uncertainty.

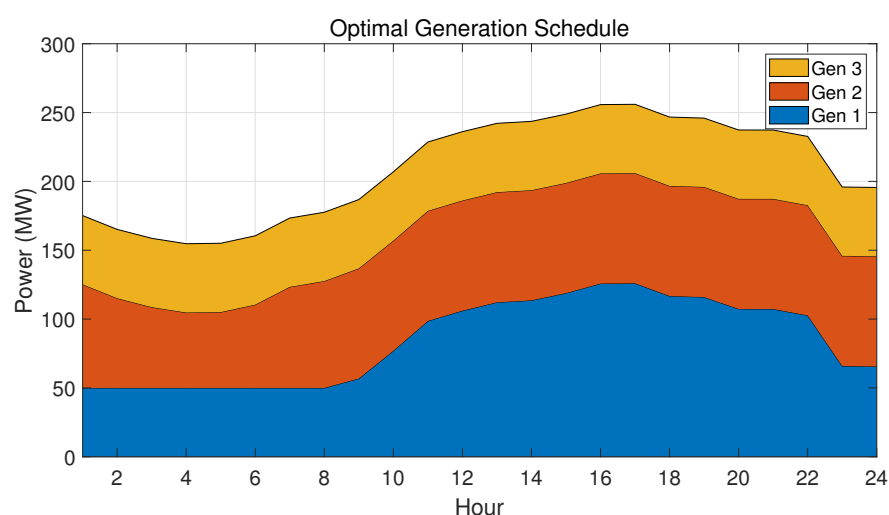


Figure 3. The optimal portfolio schedule shifted storage discharge toward high-price periods.

Table 5. Optimal renewable scheduling ($\beta = 0.95$).

Hour	Wind (MW)	Solar (MW)	Discharge (MW)	Charge (MW)	Net Position (MW)
1–6	87.3	0.0	0.0	45.2	42.1
7–9	92.1	28.4	0.0	38.7	81.8
10–14	76.4	112.3	0.0	0.0	188.7
15–18	68.9	95.6	42.3	0.0	206.8
19–22	105.2	0.0	48.6	0.0	153.8
23–24	94.6	0.0	0.0	32.1	62.5

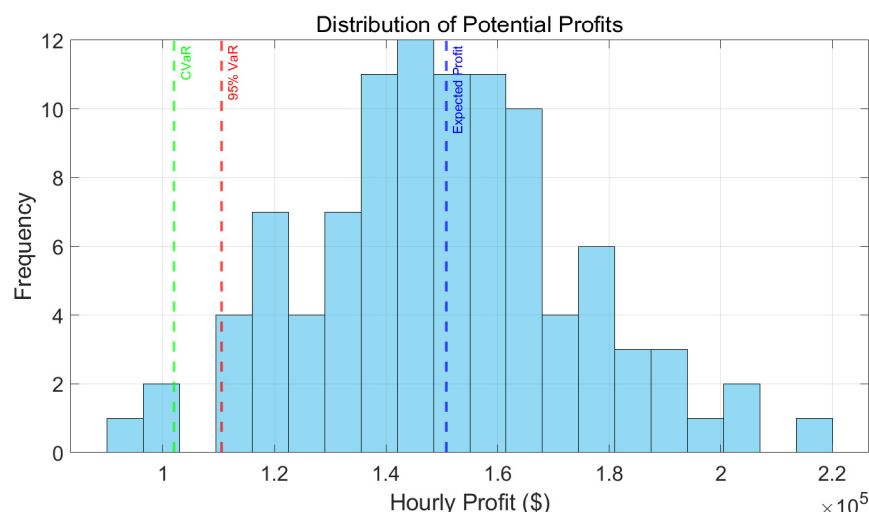


Figure 4. The profit distribution shows the expected profit, 95% VaR, and CVaR under the optimized schedule.

Figure 5 illustrates the trade-off between expected return and risk control as the risk aversion parameter β increased from 0.90 to 0.99. The expected profit decreased from \$3847/h to \$3285/h, while the worst-case profit, VaR, and CVaR improved. This pattern shows that operators can tune β to select a desired balance between average profit and downside risk protection.

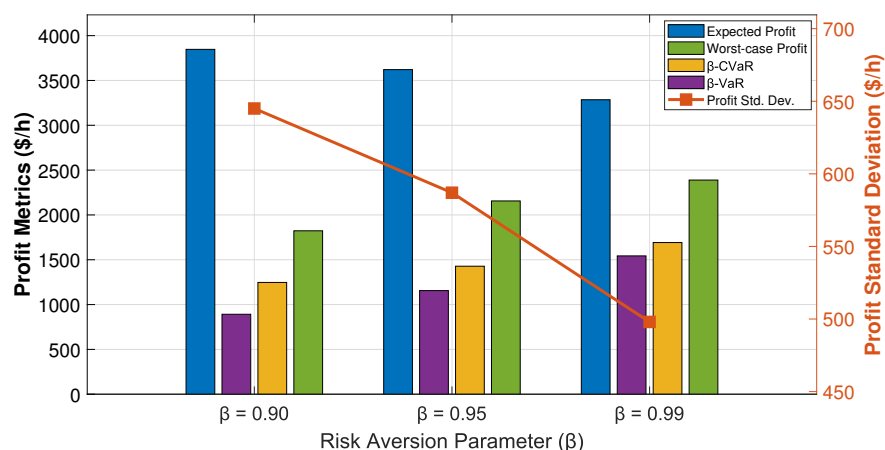


Figure 5. Increasing β improved tail risk protection at the cost of lower expected profit.

Table 6 provides the corresponding numerical comparison. As β increased from 0.90 to 0.99, the expected profit decreased by 14.6%, while the worst-case profit improved by 31%. The VaR and CVaR values also increased, confirming stronger tail risk protection under higher risk aversion.

Table 6. Risk-return analysis with T-CMRS.

Metric	$\beta = 0.90$	$\beta = 0.95$	$\beta = 0.99$
Expected Profit (\$/h)	3847	3621	3285
β -VaR (\$/h)	892	1156	1543
β -CVaR (\$/h)	1247	1428	1692
Worst-Case Profit (\$/h)	1823	2156	2389
Profit Std. Dev. (\$/h)	645	587	498

6.6. Comparison with Benchmark Methods

To contextualize the performance of T-CMRS, we compared it against four widely recognized benchmark approaches over a testing period of 50 randomly selected days. The benchmarks were a deterministic (Det) model that ignores uncertainty and uses only point forecasts; a traditional stochastic programming (SP) model that maximizes expected profit; a classic robust optimization (RO) model that optimizes against the absolute worst-case scenario within an uncertainty set; and a distributionally robust optimization (DRO) model [25] that hedges against the worst-case probability distribution within a Wasserstein ambiguity set.

The results in Table 7 highlight the balanced performance of the T-CMRS framework. Compared with the Det approach, T-CMRS showed the economic value of modeling uncertainty, achieving an 8.7% higher average profit while reducing profit volatility by 33.3%. The classic RO method, while effective at minimizing volatility, was overly conservative and led to the lowest average profits.

Table 7. Performance comparison across methods.

Method	Avg. Profit (USD/Day)	Profit Volatility	5% VaR (USD/Day)	Time (s)
Det	78,423	18,234	42,156	0.8
SP	82,156	14,523	53,234	8.5
RO	71,234	10,234	58,123	12.3
DRO	79,856	11,478	57,312	45.6
T-CMRS	85,234	12,156	61,234	8.7

In this table, Det = deterministic optimization with point forecasts only; SP = stochastic programming, maximizing expected profit; RO = robust optimization against the worst case in the uncertainty set; DRO = distributionally robust optimization with a Wasserstein ambiguity set; T-CMRS = proposed Transformer-CVaR Market Risk Scheduling framework.

The DRO method provided a middle ground between SP and RO by hedging against distributional ambiguity. It achieved lower volatility than SP (11,478 vs. 14,523) and reasonable tail risk protection (5% VaR of USD 57,312/day). However, T-CMRS outperformed DRO across all key metrics: 6.7% higher average profit, 6.9% better 5% VaR, and a substantially lower computation time (8.7 s vs. 45.6 s). The computational advantage of T-CMRS over DRO is particularly noteworthy, as DRO requires solving a semi-definite or second-order cone program, whereas T-CMRS leverages an efficient LP reformulation.

The most critical comparison is with the traditional SP model, which is designed to maximize the expected profit without specific regard for tail risk. While SP yielded a competitive average profit, it came at the cost of higher risk exposure, as indicated by its lower 5% VaR. In contrast, T-CMRS achieved a 15% improvement in the 5% VaR and a 16.3% reduction in profit volatility compared with the SP model.

This improvement in risk-adjusted performance was achieved with a small increase in computational cost (8.7 s for T-CMRS vs. 8.5 s for SP). This result shows that the CVaR-based objective, when integrated with scenarios from the transformer model, provides risk mitigation benefits without sacrificing the computational tractability required for practical scheduling applications.

6.7. Computational Performance and Scalability Analysis

Computational scalability is required for operational scheduling tools deployed in power systems of varying size and complexity. To evaluate this aspect, we tested T-CMRS

on standard test systems ranging from a 30-bus system to a 500-bus network. Table 8 summarizes the relationship between the problem size, solution time, and solver iterations.

Table 8. Scalability analysis.

Buses	Renew	Scen	Vars	Constr	Time (s)	Iters
30	3	50	288	4250	2.1	24
118	7	100	672	18,600	8.7	31
300	15	100	1440	45,300	28.4	38
500	15	150	2400	112,500	89.2	45

In this table, Buses = number of system buses; Renew = number of renewable units; Scen = number of representative scenarios; Vars = number of decision variables; Constr = number of constraints; Time = LP solver solution time; Iters = solver iterations.

As the system size increased, the number of constraints grew from approximately 4000 to over 112,000. The solution time increased from 2.1 s for the 30-bus system to 89.2 s for the 500-bus system, which remains suitable for day-ahead and rolling horizon scheduling. This computational behavior follows from the LP reformulation, which can be handled efficiently by modern interior-point solvers.

The rolling horizon T-CMRS is applicable to real-time markets because the 89.2-s solution time for the 500-bus, 150-scenario system is already suitable for day-ahead scheduling, where submission windows typically allow several minutes of computation. For real-time or intra-day markets with shorter decision windows, such as 5-min or 15-min intervals, the rolling horizon T-CMRS in Algorithm 2 provides two practical advantages. First, the rolling horizon implementation solves a shorter forward horizon, such as 4–6 h instead of 24 h, and can use a reduced scenario set, such as 50 scenarios instead of 150. This reduces the solution time to approximately 8–15 s for the 500-bus system, fitting comfortably within a 15-min dispatch interval. For 5-min balancing markets, a further scenario reduction to 20–30 scenarios and a 2-h horizon yielded solution times of approximately 2–5 s. Second, the warm start strategy in Algorithm 2 reuses the previous iteration’s optimal solution to initialize the LP solver, reducing the number of solver iterations by about 40–60%, because schedules change incrementally between consecutive rolling horizon updates.

The number of solver iterations grew from 24 for the 30-bus system to 45 for the 500-bus system, indicating that the problem structure remained tractable as the network expanded. This supports the use of T-CMRS for repeated dispatch updates as market conditions and forecasts evolve.

6.8. Sensitivity Analysis

The analysis of solution quality versus the number of scenarios in Figure 6 reveals a point of diminishing returns. Expected profit rose rapidly as the scenario count increased from a small value, but the curve began to plateau between 100 and 150 representative scenarios. This indicates that a moderate number of representative scenarios is sufficient to capture the main uncertainty effects while avoiding unnecessary computational cost.

A sensitivity analysis was performed to examine how the algorithm responded to key parameters and uncertainty ranges. Widening the forecast error bounds for all uncertainty sources by 50% resulted in a 12% reduction in expected profit, illustrating the trade-off between profitability and robustness. The scenario count analysis showed that the solution quality and expected profit stabilized once the number of representative scenarios reached 100–150. A \$10/t CO₂ increase in carbon prices increased the profitability of the renewable portfolio by 8–12%, showing the financial impact of carbon policy on scheduling decisions.

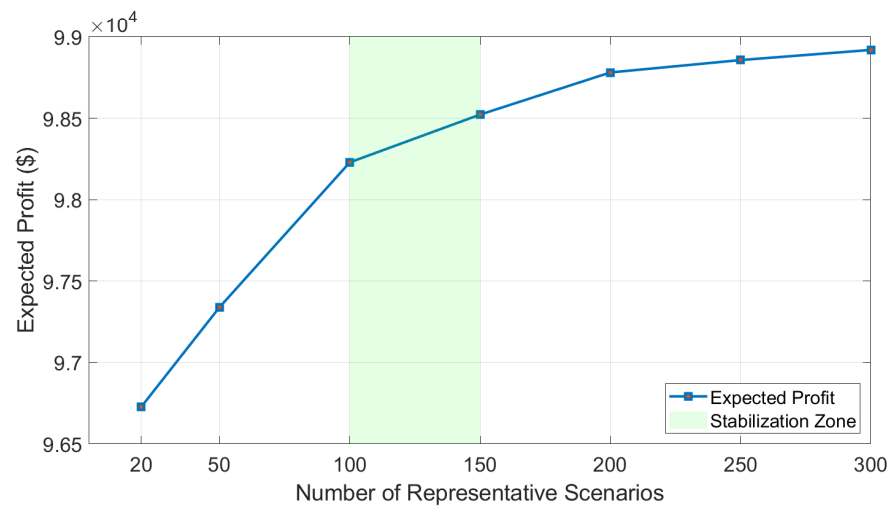


Figure 6. Solution quality stabilized after approximately 100–150 representative scenarios.

We tested the sensitivity to the fractional order α by varying the fractional order α from 0.50 to 0.95 in steps of 0.05 to evaluate its effect on scenario generation and scheduling performance. Table 9 shows that the best performance was achieved at $\alpha^* = 0.72$, consistent with the historical data estimate. More importantly, the results were robust to moderate perturbations around the optimum; changing α by ± 0.05 changed the expected profit by less than 0.3% and the 5% VaR by less than 0.5%.

Table 9. Sensitivity analysis of fractional order α .

α	Scen. MAE	Coverage	Memory KS	Exp. Profit (USD/Day)	5% VaR (USD/Day)
0.50	7.63	0.91	0.142	83,412	58,987
0.55	7.41	0.91	0.128	83,789	59,234
0.60	7.18	0.92	0.114	84,123	59,876
0.65	7.02	0.92	0.098	84,567	60,432
0.70	6.93	0.93	0.091	85,012	60,987
0.72	6.89	0.93	0.087	85,234	61,234
0.75	6.91	0.93	0.089	85,187	61,178
0.80	7.05	0.92	0.102	84,876	60,765
0.85	7.23	0.92	0.118	84,432	60,321
0.90	7.48	0.91	0.134	83,987	59,876
0.95	7.82	0.91	0.156	83,456	59,234

The performance degradation became more pronounced when $\alpha < 0.60$ (insufficient memory) or $\alpha > 0.90$ (excessive memory), both of which deviated substantially from the empirically estimated value. This confirms that the proposed model is not overly sensitive to small estimation errors in α while still benefiting from a data-consistent fractional order.

We further evaluated CVaR performance under extreme carbon policy shocks by considering three tail risk carbon policy scenarios, namely (1) a 30% overnight carbon price increase, (2) a 50% carbon price spike, and (3) a 40% carbon price drop. Under scenario (1), expected profit increased by 22–28% because carbon-free renewable generation became more valuable, while the 5% VaR decreased by about 8% due to wider downside exposure. Under scenario (2), expected profit increased by 35–42%, but the CVaR degraded by about 12% relative to the baseline because the extreme spike partially exceeded the calibrated

scenario range. Under scenario (3), the expected profit decreases by 15–18%, and the CVaR decreased by about 20%, as lost carbon revenue directly affected the worst-profit scenarios. The CVaR formulation mitigated these shocks by explicitly optimizing the lower-tail profit distribution, and the copula-based scenario generation helped represent the tail dependence between electricity and carbon prices. The rolling horizon implementation further allows rapid re-optimization when the carbon-price distribution shifts. However, truly unprecedented regime shifts may require scenario-set expansion or distributionally robust extensions to maintain adequate tail risk protection.

7. Conclusions

This paper proposed a risk-averse scheduling framework, T-CMRS, that integrates transformer-based prediction with fractional-order uncertainty modeling and CVaR optimization to enhance the market performance of renewable energy under multi-dimensional uncertainty. The framework uses a transformer model to improve price prediction accuracy, employs Caputo fractional derivatives to capture the memory effects inherent in electricity price dynamics, and utilizes a CVaR-based method to manage tail risk, while an LP reformulation ensures computational efficiency. Case studies show that the transformer model reduced the price prediction error by 21.7% compared with LSTM, the fractional-order uncertainty modeling reduced the scenario generation error by 12.3% compared with integer-order approaches, and the integrated framework increased average profits by 8.7% with 33.3% lower volatility compared with deterministic optimization. These results support the value of combining probabilistic forecasting with fractional-order memory-aware modeling and risk-averse optimization for renewable energy scheduling in coupled electricity and carbon markets.

The key figures and tables further clarify the contribution of each component. Figure 1 and Table 3 demonstrate that the transformer provided the most accurate and best-calibrated forecasting input for risk scheduling. Table 4 validates the value of fractional-order memory modeling by showing a lower scenario error and stronger memory-structure preservation than integer-order alternatives. Figure 5 and Table 6 illustrate how the CVaR parameter β controlled the trade-off between expected profit and tail-risk protection. Table 7 confirms that T-CMRS outperformed deterministic, stochastic, robust, and distributionally robust benchmarks in risk-adjusted performance. Table 8 verifies scalability to 500-bus systems, while Figure 6 shows that 100–150 representative scenarios were sufficient for stable solution quality. Together, these results demonstrate the practical value of coupling probabilistic forecasting, fractional-order memory modeling, and CVaR-based optimization.

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

Funding: This research was funded by the Shanghai Science and Technology Innovation Action Plan Morning Star Project (Sail Special), grant number 22YF1411900.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors would like to thank the anonymous reviewers for their constructive comments.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

T-CMRS	Transformer-CVaR Market Risk Scheduling
CVaR	Conditional value at risk
VaR	Value at risk
LMP	Locational marginal price
LP	Linear program
SP	Stochastic programming
RO	Robust optimization
DRO	Distributionally robust optimization
LSTM	Long short-term memory
GRU	Gated recurrent unit
ARIMA	AutoRegressive Integrated Moving Average
PJM	Pennsylvania-New Jersey-Maryland Interconnection
EUA	European Union Allowance
ETS	Emissions trading system
NREL	National Renewable Energy Laboratory
MAE	Mean absolute error
RMSE	Root mean square error
MAPE	Mean absolute percentage error
PI	Prediction interval
FO	Fractional order
Caputo	Caputo fractional derivative

References

1. Yan, N.; Ma, G.; Li, X.; Guerrero, J.M. Low-Carbon Economic Dispatch Method for Integrated Energy System Considering Seasonal Carbon Flow Dynamic Balance. *IEEE Trans. Sustain. Energy* **2023**, *14*, 576–586. [\[CrossRef\]](#)
2. Wang, S.; Hui, H.; Zhai, J. Short-Term Reliability Assessment of Integrated Power-Gas Systems with Hydrogen Injections Using Universal Generating Function. *IEEE Trans. Ind. Appl.* **2023**, *59*, 5760–5773. [\[CrossRef\]](#)
3. Peng, F.Z.; Liu, C.C.; Li, Y.; Jain, A.K.; Vinnikov, D. Envisioning the Future Renewable and Resilient Energy Grids—A Power Grid Revolution Enabled by Renewables, Energy Storage, and Energy Electronics. *IEEE J. Emerg. Sel. Top. Ind. Electron.* **2024**, *5*, 8–26. [\[CrossRef\]](#)
4. Zhou, W.; Lai, J.; Li, X.; Zeng, Z. Dynamic Vertical Peer-to-Peer Carbon Right Sharing Management for Microgrid Clusters with Multiple Structural Demand Response. *IEEE Trans. Ind. Inform.* **2025**, *21*, 6038–6048. [\[CrossRef\]](#)
5. Podlubny, I. *Fractional Differential Equations*; Academic Press: San Diego, CA, USA, 1999.
6. Kilbas, A.A.; Srivastava, H.M.; Trujillo, J.J. *Theory and Applications of Fractional Differential Equations*; Elsevier: Amsterdam, The Netherlands, 2006.
7. Wang, C.; Wang, Y.; Ding, Z.; Zheng, T.; Hu, J.; Zhang, K. A Transformer-Based Method of Multienergy Load Forecasting in Integrated Energy System. *IEEE Trans. Smart Grid* **2022**, *13*, 2703–2714. [\[CrossRef\]](#)
8. Zhu, Q.; Chang, P.; Li, S.; Xiao, Y.; Wu, T.; Li, L.; Saeed, A.; Li, C. Applications of Equilibrium Decision Theory in Low-Carbon Economic Operation of Wind-Solar-Hydrogen Integrated Energy System Considering Prediction Uncertainties. *IEEE Trans. Ind. Inform.* **2025**, *21*, 6230–6239. [\[CrossRef\]](#)
9. Xu, L.; Zeng, H.; Lin, N.; Yang, Y.; Guo, Q.; Poor, H.V. Entropic Value-at-Risk Constrained Optimal Power Flow Considering Distribution Network Outages During Extreme Events. *IEEE Trans. Power Syst.* **2025**, *40*, 1184–1187. [\[CrossRef\]](#)
10. Ran, X.; Tay, W.P.; Lee, C. High-Performance Optimization Model Based on Novel Conditional Value At Risk Metric for Power Grids with High Wind Power Penetration. *IEEE Trans. Ind. Inform.* **2025**, *21*, 5689–5700. [\[CrossRef\]](#)
11. Zhu, Z.; Chan, K.W.; Bu, S.; Zhou, B.; Xia, S. Nash Equilibrium Estimation and Analysis in Joint Peer-to-Peer Electricity and Carbon Emission Auction Market with Microgrid Prosumers. *IEEE Trans. Power Syst.* **2023**, *38*, 5768–5780. [\[CrossRef\]](#)

12. Baringo, L.; Conejo, A.J. Offering Strategy of Wind-Power Producer: A Multi-Stage Risk-Constrained Approach. *IEEE Trans. Power Syst.* **2016**, *31*, 1420–1429. [\[CrossRef\]](#)
13. Li, Z.; Wu, W.; Zhang, B.; Wang, B. Decentralized Multi-Area Dynamic Economic Dispatch Using Modified Generalized Benders Decomposition. *IEEE Trans. Power Syst.* **2016**, *31*, 526–538. [\[CrossRef\]](#)
14. Nourollahi, R.; Gholizadeh-Roshanagh, R.; Feizi-Aghakandi, H.; Zare, K.; Mohammadi-Ivatloo, B. Power Distribution Expansion Planning in the Presence of Wholesale Multimarkets. *IEEE Syst. J.* **2023**, *17*, 1684–1693. [\[CrossRef\]](#)
15. Rashidzadeh-Kermani, H.; Vahedipour-Dahraie, M.; Shafie-Khah, M.; Siano, P. A Regret-Based Stochastic Bi-Level Framework for Scheduling of DR Aggregator Under Uncertainties. *IEEE Trans. Smart Grid* **2020**, *11*, 3171–3184. [\[CrossRef\]](#)
16. Jiang, W.; Liu, J.; Shafqat, J.; Lei, Y. CVaR-Based Reserve Optimization for Isolated Multi-Energy Complementary Generation System in Mixed Time Scale. *CSEE J. Power Energy Syst.* **2024**, *10*, 2597–2609. [\[CrossRef\]](#)
17. Zou, Y.; Xu, Y.; Zhang, C. A Risk-Averse Adaptive Stochastic Optimization Method for Transactive Energy Management of a Multi-Energy Microgrid. *IEEE Trans. Sustain. Energy* **2023**, *14*, 1599–1611. [\[CrossRef\]](#)
18. Tu, Q.; Miao, S.; Yao, F.; Li, Y.; Yin, H.; Han, J.; Zhang, D.; Yang, W. Forecasting Scenario Generation for Multiple Wind Farms Considering Time-series Characteristics and Spatial-temporal Correlation. *J. Mod. Power Syst. Clean Energy* **2021**, *9*, 837–848. [\[CrossRef\]](#)
19. Li, Y.; Li, C.; Chen, G.; Zhou, X.; Dong, Z. Multi-Task Graph Adaptive Learning for Multivariate Electricity Price Short-Term Forecasting in Australia's National Electricity Market. *IEEE Trans. Power Syst.* **2025**, *40*, 530–542. [\[CrossRef\]](#)
20. Liu, S.; Jiang, Y.; Lin, Z.; Wen, F.; Ding, Y.; Yang, L. Data-driven Two-step Day-ahead Electricity Price Forecasting Considering Price Spikes. *J. Mod. Power Syst. Clean Energy* **2023**, *11*, 523–533. [\[CrossRef\]](#)
21. Casagrande, V.; Ferienc, M.; Rodrigues, M.R.D.; Boem, F. Online End-to-End Learning-Based Predictive Control for Microgrid Energy Management. *IEEE Trans. Control Syst. Technol.* **2025**, *33*, 463–478. [\[CrossRef\]](#)
22. Kim, H.J.; Kim, M.K. Real-Time Price Forecasting Combined Deep Reinforcement Learning for Predictive Home Energy Management System. *IEEE Internet Things J.* **2025**, *12*, 34806–34821. [\[CrossRef\]](#)
23. Bottieau, J.; Wang, Y.; De Grève, Z.; Vallée, F.; Toubreau, J.F. Interpretable Transformer Model for Capturing Regime Switching Effects of Real-Time Electricity Prices. *IEEE Trans. Power Syst.* **2023**, *38*, 2162–2176. [\[CrossRef\]](#)
24. Yuan, H.; Xu, Y. Optimal Preventive Generation Rescheduling for Improving Small- and Large-Disturbance Rotor Angle Stabilities of Power Systems Considering Wind Power Uncertainty. *J. Mod. Power Syst. Clean Energy* **2024**, *13*, 1568–1579. [\[CrossRef\]](#)
25. Zhou, A.; Yang, M.; Fang, X.; Zhang, Y. Addressing Wind Power Forecast Errors in Day-Ahead Pricing with Energy Storage Systems: A Distributionally Robust Joint Chance-Constrained Approach. *IEEE Trans. Sustain. Energy* **2024**, *15*, 1754–1767. [\[CrossRef\]](#)
26. Xuan, A.; Shen, X.; Guo, Q.; Sun, H. Two-Stage Planning for Electricity-Gas Coupled Integrated Energy System with Carbon Capture, Utilization, and Storage Considering Carbon Tax and Price Uncertainties. *IEEE Trans. Power Syst.* **2023**, *38*, 2553–2565. [\[CrossRef\]](#)
27. Xie, Y.; Xu, Y.; Yi, Z.; Lin, S.; Zhang, B.; Zhu, X.; Rong, S. A Hybrid Stochastic-Robust Planning Approach for the Flexible Devices in an Islanded Integrated Energy System Considering Multiple Uncertainties and Demand Response. *IEEE Trans. Ind. Appl.* **2025**, *61*, 8037–8050. [\[CrossRef\]](#)
28. Zhang, J.; Chang, X.; Xue, Y.; Bai, X.; Li, Z.; Wang, P.; Sun, H. Optimal Planning for Electricity-Gas-Hydrogen Integrated Energy Systems Considering Intertemporal Long-term Hydrogen Storage and Multiple Uncertainties. *IEEE Trans. Power Syst.* **2025**, *40*, 4660–4674. [\[CrossRef\]](#)
29. Baleanu, D.; Diethelm, K.; Scalas, E.; Trujillo, J.J. *Fractional Calculus: Models and Numerical Methods*, 2nd ed.; World Scientific: Singapore, 2016. [\[CrossRef\]](#)
30. Houllier, M.A.; de Menezes, L.M. A fractional cointegration analysis of European electricity spot prices. In Proceedings of the 2012 9th International Conference on the European Energy Market, Florence, Italy, 10–12 May 2012; pp. 1–6. [\[CrossRef\]](#)
31. Mainardi, F.; Gorenflo, R. On Mittag-Leffler-type functions in fractional evolution processes. *J. Comput. Appl. Math.* **2000**, *118*, 283–299. [\[CrossRef\]](#)
32. Gorenflo, R.; Kilbas, A.A.; Mainardi, F.; Rogosin, S.V. *Mittag-Leffler Functions, Related Topics and Applications*; Springer: Berlin, Germany, 2014. [\[CrossRef\]](#)
33. Li, C.; Zeng, F. *Numerical Methods for Fractional Calculus*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2015. [\[CrossRef\]](#)
34. Diethelm, K.; Ford, N.J.; Freed, A.D. A predictor-corrector approach for the numerical solution of fractional differential equations. *Nonlinear Dyn.* **2002**, *29*, 3–22. [\[CrossRef\]](#)
35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
36. Worth, K.; Plamus, T.; Majak, J. Design optimization of nonwovens made from mechanically recycled textile fibres. *Compos. Struct.* **2026**, *387*, 120330. [\[CrossRef\]](#)
37. Nelsen, R.B. *An Introduction to Copulas*, 2nd ed.; Springer: New York, NY, USA, 2006. [\[CrossRef\]](#)

38. Dupacova, J.; Growe-Kuska, N.; Romisch, W. Scenario reduction in stochastic programming: An approach using probability metrics. *Math. Program.* **2003**, *95*, 493–511. [[CrossRef](#)]
39. Simonsen, I. Measuring anti-correlations in the Nordic electricity spot market by wavelets. *Phys. A Stat. Mech. Its Appl.* **2003**, *322*, 597–606. [[CrossRef](#)]
40. Weron, R. Estimating long-range dependence: Finite sample properties and confidence intervals. *Phys. A Stat. Mech. Its Appl.* **2002**, *312*, 285–299. [[CrossRef](#)]
41. Botterud, A.; Zhou, Z.; Wang, J.; Bessa, R.J.; Keko, H.; Sumaili, J.; Miranda, V. Wind power trading under uncertainty in LMP markets. *IEEE Trans. Power Syst.* **2012**, *27*, 894–903. [[CrossRef](#)]
42. Gneiting, T.; Katzfuss, M. Probabilistic forecasting. *Annu. Rev. Stat. Its Appl.* **2014**, *1*, 125–151. [[CrossRef](#)]
43. Contreras-Reyes, J.E.; Palma, W. Statistical analysis of autoregressive fractionally integrated moving average models in R. *Comput. Stat.* **2013**, *28*, 2309–2331. [[CrossRef](#)]

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