

Evaluating Intervention Effectiveness during Extreme Price Events: A Counterfactual Analysis of Romania's 2024 Energy Crisis

Alexandru-Victor ANDREI*

Bucharest University of Economic Studies, Romania

**Corresponding author, andreilvictor23@stud.ase.ro*

Daniel-Traian PELE

IDA Institute Digital Assets Bucharest University of Economic Studies

Institute for Economic Forecasting, Romanian Academy, Bucharest, Romania

danpele@ase.ro

Alexandru-Adrian CRAMER

Bucharest University of Economic Studies, Romania

crameralexandru12@stud.ase.ro

Abstract. *Extreme electricity price events pose significant challenges for energy policymakers, yet the question of which interventions most effectively mitigate such crises remains under-explored. This paper develops a counterfactual analysis framework to evaluate intervention effectiveness during Romania's June–July 2024 electricity price crisis, when day-ahead prices reached 1,022 euros per megawatt-hour. Using a dataset of 59,160 hourly observations from the European Network of Transmission System Operators for Electricity Transparency Platform, we trained a Random Forest model on 70 engineered features and generated 147 counterfactual scenarios across three intervention categories: renewable generation increase, demand reduction, and combined approaches. To address the fundamental limitation that tree-based models underpredict extreme events by 68 percent, we validated all findings through cross-model comparison with Ridge regression, achieving a Spearman rank correlation of 0.73 between intervention rankings across both architectures. Results reveal that demand response of 10 percent load reduction is the robust first-tier intervention, achieving 70 percent price reduction with unanimous cross-model consensus and full operational feasibility. A critical finding is that the peak event occurred during nighttime hours when solar generation stood at 0.2 percent of capacity, rendering solar-based interventions physically impossible and identifying battery energy storage as critical enabling infrastructure. No feasible intervention achieves normal price levels, reflecting the genuine severity of the crisis. All scenarios were assessed against operational constraints grounded in eight academic references covering forecasting accuracy limits, demand response capacity, and activation warning times. The paper contributes a methodological framework for counterfactual analysis under model uncertainty, demonstrating that meaningful policy guidance can be derived from relative effectiveness rankings even when absolute predictions are unreliable.*

Keywords: explainable artificial intelligence, electricity price forecasting, early warning system, load-generation balance, machine learning interpretability, Romanian energy market, price spike detection.

Introduction

In the summer of 2024, electricity prices across Southeastern Europe surged to levels not seen even during the worst of the 2021–2022 energy crisis. Romania's day-ahead market recorded hourly prices exceeding 1,000 euros per megawatt-hour, driven by a convergence of extreme heat, reduced hydropower output due to drought, limited cross-border transmission capacity, and increased

DOI: 10.2478/picbe-2026-0013

© 2026 A.-V. Andrei; A.-A. Cramer; D. Diaconu, published by De Gruyter.
This work is licensed under the Creative Commons Attribution 4.0 License.

electricity exports to Ukraine (Balkan Green Energy News, 2024; Romania Insider, 2024a). Greece, Bulgaria, and Romania subsequently announced a joint initiative to address what they described as systematic isolation of Southeastern Europe from the broader European electricity market (Romania Insider, 2024b). These events underscore a growing challenge for energy policymakers: as electricity systems become more dependent on variable renewable generation, they also become more vulnerable to extreme price events that exceed historical norms and resist conventional forecasting. Recent research has confirmed that Romanian electricity prices have been fundamentally altered by successive disruptions, with traditional forecasting approaches losing predictive power during the pandemic and the Ukraine conflict (Bara & Oprea, 2024a; Bara et al., 2023).

Understanding how to respond to such crises requires not only knowing why prices spiked — a question addressed by attribution-based explainability methods such as Shapley additive explanations — but also knowing what could have been done differently. This second question is the domain of counterfactual analysis: by systematically modifying the conditions that prevailed during a crisis event and observing how model predictions change, we can rank the relative effectiveness of different interventions and identify which responses offer the greatest potential for mitigation.

However, applying counterfactual analysis to extreme price events presents a methodological challenge that has received limited attention in the literature. Machine learning models, particularly tree-based algorithms, are trained predominantly on normal market conditions and cannot reliably predict prices during unprecedented events that fall outside the training distribution. This creates a paradox: the situations where counterfactual guidance is most needed — extreme crises — are precisely the situations where prediction models are least reliable.

This paper addresses this paradox by developing a counterfactual framework that focuses on relative effectiveness rankings rather than absolute price predictions, validates findings across fundamentally different model architectures (a Random Forest and a Ridge regression), and constrains scenarios to operationally feasible ranges grounded in academic literature. We apply this framework to Romania's June–July 2024 electricity price crisis and address five research questions: (1) which intervention types are most effective at reducing extreme prices; (2) what is the relative impact of different intervention magnitudes; (3) how do interventions combine to produce synergistic effects; (4) which interventions are feasible within operational constraints; and (5) what time-of-day and weather constraints affect intervention effectiveness.

Literature review

Extreme price events in electricity markets

Electricity prices exhibit occasional extreme spikes that deviate sharply from normal levels, driven by the convergence of high demand, low renewable output, transmission constraints, and fuel price shocks (Weron, 2014). In 2024, Southeastern European markets experienced particularly severe episodes: Romania's OPCOM exchange recorded hourly prices reaching 700 euros per megawatt-hour in mid-July, while Hungary's HUPX exchange saw prices exceeding 1,000 euros (Balkan Green Energy News, 2024). Analysis of these events revealed that they were not primarily caused by demand increases — regional consumption was comparable to the 2021 heatwave — but rather by a reduction in effective cross-border transmission capacity and the closure of 2,000 megawatts of coal-fired capacity in the region since 2021 (Balkan Green Energy News, 2024). Cincinelli and Pellini (2025) confirmed through econometric analysis of 12 European wholesale markets that geopolitical events, particularly disruptions to natural gas supplies, are the primary drivers of

extreme price bubbles, with a one-point increase in geopolitical risk raising bubble probability by 8.23 percent. The European price wall phenomenon, in which Southeastern European prices diverge sharply from Western European levels due to infrastructure limitations and differing market coupling methodologies, has been documented as a persistent structural vulnerability (CEEnergy News, 2025). These findings mean that extreme price events are likely to recur, making the study of effective crisis responses increasingly urgent.

Forecasting electricity prices during such extreme events remains a major challenge. Borges, Maia, and Guerreiro (2024) demonstrated that standard forecasting models fail during crisis periods, proposing a hybrid LSTM-ARIMA architecture specifically designed for extreme event prediction. In the Romanian context, Bara and Oprea (2024a) showed that the shape of the electricity price curve has been fundamentally altered by recent disruptions, with previously influential factors losing predictive power and new variables emerging. Their work on the Romanian day-ahead market demonstrated that adjusting training periods according to price volatility—shortening the window during high-growth periods—can improve forecast accuracy by up to 48 percent. Bara, Oprea, and Baroiu (2023) further confirmed the difficulty of Romanian electricity price forecasting during the pandemic and the Ukraine conflict using LSTM architectures, while Bara and Oprea (2024b) showed that price fluctuations in one East-European country can be substantially explained by previous-day prices in neighbouring markets, reflecting the interconnected nature of regional crisis transmission.

Explainable artificial intelligence in energy systems

The application of explainable artificial intelligence to energy and power systems has grown rapidly since 2020, with particular focus on making machine learning predictions interpretable for grid operators and market participants (Machlev et al., 2022). The most commonly applied techniques include Shapley additive explanations for feature attribution and local interpretable model-agnostic explanations for instance-level interpretation. In energy price forecasting specifically, Pesenti and O’Sullivan (2025) used deep neural networks combined with SHAP and gradient-based saliency maps to analyse price dynamics across five European electricity markets, introducing the novel concepts of SSHAP values and SSHAP lines for high-dimensional tabular models. Tschora et al. (2022) applied SHAP-based analysis to day-ahead price forecasting and showed that incorporating price histories from neighbouring countries significantly improves forecast quality, demonstrating the value of explainability for understanding cross-border market dependencies.

A critical challenge identified in recent work is the trustworthiness of model predictions under changing market conditions. Heistrene et al. (2023) proposed an explainability-based trust algorithm that generates trust scores reflecting prediction quality for each new input, achieving over 85 percent accuracy in identifying reliable predictions under stable conditions. Crucially, under distribution shift—precisely the condition that characterises extreme price events—the algorithm showed equivalent accuracy in identifying unreliable predictions. This finding directly supports our paper’s approach of focusing on relative rankings rather than absolute predictions during out-of-distribution events. Melgar-Garcia and Troncoso (2024) developed a real-time explainable forecasting algorithm for Spanish electricity prices that detects novelties and unexpected behaviours in time series streams, further illustrating the growing integration of explainability with anomaly detection in energy markets.

However, the predominant focus in the literature has been on attribution—explaining why a prediction was made—rather than on prescription—identifying what actions could change an

outcome. This distinction mirrors the broader classification of explainable artificial intelligence methods into attribution-based approaches (such as Shapley values) and example-based approaches (such as counterfactual explanations) (Molnar, 2020).

Counterfactual analysis for policy evaluation

Counterfactual explanations were formally introduced to the machine learning literature by Wachter, Mittelstadt, and Russell (2017), who proposed them as a means of explaining automated decisions without requiring access to model internals. Since then, a substantial body of work has developed, with recent surveys cataloguing over 100 methods for generating counterfactual explanations (Verma et al., 2024). Most counterfactual research has focused on classification problems — for example, identifying what a loan applicant would need to change to receive approval. Applications to regression problems, where the output is a continuous value rather than a class label, remain less explored. Furthermore, existing methods typically assume that the prediction model is reliable across the input space, an assumption that breaks down when analysing extreme events that lie outside the training distribution. Our paper addresses both of these gaps by applying counterfactual analysis to a regression target (electricity price) under conditions of acknowledged model uncertainty, and by introducing cross-model validation as a robustness check for counterfactual rankings.

Intervention strategies for electricity markets

The literature on crisis intervention in electricity markets spans demand response, renewable integration, and storage deployment. Gils (2014) estimated that the technical demand response potential in Europe ranges from 15 to 30 percent of peak load, but the operational potential — accounting for contractual, economic, and social constraints — is only 5 to 12 percent. The European Network of Transmission System Operators for Electricity (2020) documented that emergency demand response programmes across European transmission system operators can mobilise 3 to 8 percent of peak load within two to four hours. For Romania specifically, the transmission system operator Transelectrica maintains approximately 800 megawatts of contracted demand response capacity, representing about 11.4 percent of peak load during the summer 2024 crisis (Transelectrica, 2023).

On the renewable forecasting side, Inman, Pedro, and Coimbra (2013) showed that day-ahead solar forecast errors typically range from 20 to 40 percent, while Giebel et al. (2011) documented wind forecast errors of 10 to 25 percent of installed capacity. Ganczarek-Gamrot et al. (2025) provided recent evidence from seven European bidding zones showing that zones dominated by low-controllability renewable generation exhibit greater price volatility and reduced forecast accuracy, while zones with higher shares of controllable sources demonstrate more stable prices — a finding that directly supports our focus on demand response as a controllable intervention. Battery energy storage has emerged as a critical enabling technology, particularly for addressing the evening peak problem in systems with high solar penetration, where generation drops to near zero precisely when demand peaks (European Commission, Directorate-General for Energy, 2021).

The present paper contributes to this literature by systematically evaluating these intervention categories within a unified counterfactual framework, validating findings across model architectures, and grounding all scenarios in literature-based feasibility constraints specific to the Romanian context.

Methodology

Data sources and preparation

The empirical analysis draws on multiple open-access data sources covering the Romanian electricity market over the period January 2018 to mid-2024. The primary dataset was extracted from the European Network of Transmission System Operators for Electricity Transparency Platform, which provides hourly records of electricity generation by fuel type (fossil gas, nuclear, solar, wind onshore, hydro run-of-river, and hydro water reservoir), day-ahead generation forecasts for solar and wind, day-ahead total load forecasts, and actual total load. Day-ahead wholesale electricity prices, expressed in euros per megawatt-hour, were obtained from the European Wholesale Electricity Price Data repository, which aggregates and cleans price information originally published by the same transparency platform. These are daily day-ahead prices, meaning that every hour within a given calendar day carries the same price value.

Temperature data were collected through the Open-Meteo Historical Weather Application Programming Interface for ten Romanian cities representing distinct climatic zones: Bucharest, Timisoara, Iasi, Cluj-Napoca, Constanta, Brasov, Craiova, Suceava, Deva, and Galati. Each extraction used latitude and longitude coordinates for the respective city, with the European Central timezone setting applied across all locations. Commodity prices for crude oil, natural gas, and carbon dioxide emission allowances were sourced from Refinitiv and are reported as daily values, so that each hour of the same trading day receives the same commodity price.

The raw dataset comprises 59,160 hourly observations across 47 variables. To manage computational cost — counterfactual generation and cross-model validation require many thousands of model evaluations — the analysis uses the most recent 5,000 observations (approximately 208 days), selected chronologically from the end of the dataset using a tail-selection approach. This strategy preserves the temporal ordering of the data and avoids any look-ahead bias that random sampling would introduce. Importantly, all feature engineering (described below) was performed on the full dataset before the subsample was extracted, ensuring that rolling-window statistics and lagged variables have sufficient historical depth and do not suffer from edge effects. The subsample size of 5,000 observations was chosen to balance computational tractability — counterfactual generation and cross-model validation require evaluating thousands of model predictions per scenario — with sufficient temporal coverage of the crisis period. While training on the full 59,160 observations could potentially improve model accuracy for normal conditions, the primary purpose of the model in this paper is relative scenario comparison rather than absolute forecast precision.

The 5,000 observations were split chronologically into a training set of 4,000 observations (80 percent) and a test set of 1,000 observations (20 percent). The test set covers the June–July 2024 period, which includes the extreme price crisis that is the focus of this paper.

Feature engineering

Starting from the 47 raw variables, we constructed 70 features organised into six categories designed to capture distinct aspects of electricity market dynamics.

The first category comprises cyclical time encodings. Because time-of-day, day-of-week, and month-of-year are circular variables (hour 23 is close to hour 0, December is close to January), we encoded each using sine and cosine transformations. This gives the model a smooth, continuous representation of temporal patterns rather than treating these variables as arbitrary integers. The cyclical encoding of a temporal variable t with period T is defined as:

$$t_{sin} = \sin\left(\frac{2\pi t}{T}\right) \quad (1)$$

$$t_{cos} = \cos\left(\frac{2\pi t}{T}\right) \quad (2)$$

where $T = 24$ for hour-of-day, $T = 7$ for day-of-week, and $T = 12$ for month-of-year. This pair of transformations maps each temporal value onto a unit circle, ensuring that the model treats adjacent time points as numerically close even when they span the boundary of the period (for example, hour 23 and hour 0).

The second category includes price lags at multiple horizons: one hour, six hours, twelve hours, one day, two days, three days, seven days, and fourteen days prior. These allow the model to capture short-term momentum (prices tend to persist over consecutive hours), weekly seasonality (similar prices on the same day of successive weeks), and medium-term trends.

The third category consists of interaction features that combine two or more raw variables to represent domain-relevant relationships. The most important of these is the residual load forecast, defined as the day-ahead total load forecast minus day-ahead solar and wind generation forecasts. Residual load measures the portion of demand that must be met by dispatchable generation (fossil gas, nuclear, hydro), and is widely recognised in the energy economics literature as a key driver of wholesale prices. We also computed the renewable share forecast (the ratio of forecasted solar plus wind generation to forecasted total load), a load-temperature interaction term (the product of load forecast and lagged temperature, capturing heat-wave-driven demand surges), a peak-hour load interaction (load forecast multiplied by a binary indicator for peak hours 8–11 and 18–21), and a gas cost proxy (lagged gas price multiplied by lagged fossil gas generation).

Residual load, the key derived feature, is computed as:

$$L_{res} = L_{forecast} - G_{solar}^{forecast} - G_{wind}^{forecast} \quad (3)$$

where $L_{forecast}$ is the day-ahead total load forecast in megawatts and $G_{solar}^{forecast}$ and $G_{wind}^{forecast}$ are the day-ahead solar and wind generation forecasts respectively.

The fourth category captures rolling statistics computed over seven-day and thirty-day windows: price volatility (standard deviation), price moving averages, load variability, and solar and wind forecast variability. These features help the model distinguish between calm and turbulent market periods.

The fifth category addresses forecast errors. We computed the difference between day-ahead forecasts and actual outcomes for both solar and wind generation at the one-day lag, as well as seven-day rolling averages of these errors (forecast bias). These features allow the model to learn whether systematic over- or under-forecasting of renewables is contributing to price stress.

The sixth and final category includes domain-specific features: hydro reservoir share (the proportion of total hydro generation coming from reservoirs, as opposed to run-of-river, which signals the availability of flexible storage-backed generation), nuclear availability (a binary indicator for whether nuclear units were operating), total generation, supply-demand balance (total generation minus total load), and the renewable generation share.

All features with missing values — primarily those arising from rolling windows at the start of the dataset or from lagged variables at the boundary — were filled using forward-fill, backward-fill, and zero-imputation in that order of priority. Infinite values were replaced with zero.

Prediction model

We trained two candidate models on the 4,000-observation training set: a Random Forest regressor and a gradient-boosted tree model. The Random Forest was configured with 200 trees, a maximum depth of 20, a minimum of 10 samples required to split an internal node, a minimum of 4 samples per leaf, and the square root of the total number of features considered at each split. The gradient-boosted model used 200 boosting rounds, a maximum depth of 8, a learning rate of 0.05, row subsampling at 80 percent, and column subsampling at 80 percent, with L_1 and L_2 regularisation.

A Random Forest constructs an ensemble of B decision trees, each trained on a bootstrap sample of the data. The prediction for a new observation x is the average of the individual tree predictions:

$$\widehat{y}_{\text{RF}}(x) = \frac{1}{B} \sum_{b=1}^B h_b(x) \quad (4)$$

where $h_b(x)$ is the prediction of the b -th tree.

On the held-out test set, the Random Forest achieved a coefficient of determination (R^2) of 0.5963 and the gradient-boosted model achieved 0.5699. We therefore selected the Random Forest as the primary model for counterfactual generation. While this R^2 is moderate by forecasting standards, it is important to note that the model serves as a scenario evaluation tool in this paper — its purpose is not to predict prices with pinpoint accuracy, but to provide a consistent framework for comparing the relative effectiveness of different interventions.

Model limitations and validity for counterfactual analysis

Before generating counterfactual scenarios, we assessed the model's ability to predict prices across different ranges. This step is critical because counterfactual analysis relies on the model's predictions to evaluate intervention outcomes, and any systematic bias must be understood and communicated transparently.

Model performance was evaluated using the coefficient of determination (R^2), mean absolute error (MAE), and a range-specific underprediction metric:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \widehat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \widehat{y}_i| \quad (6)$$

$$\text{Underprediction}_k = \frac{1}{n_k} \sum_{i \in S_k} \frac{y_i - \widehat{y}_i}{y_i} \times 100 \quad (7)$$

where y_i is the actual price, $\widehat{y}_i$ is the predicted price, $\bar{y}$ is the mean actual price, and S_k denotes the subset of observations in price range k with n_k observations.

The analysis revealed a clear pattern: prediction accuracy deteriorates as prices increase. In the 100–200 euros per megawatt-hour range, the mean absolute error is 18.97 euros per megawatt-

hour and the average underprediction is only 0.3 percent. In the 300–500 range, mean absolute error rises to 122.56 and underprediction reaches 29.9 percent. For prices above 500 euros per megawatt-hour, the mean absolute error is 362.59 and underprediction reaches 53.9 percent. For the specific subset of extreme events above the 95th percentile (50 observations, mean actual price 513.84 euros per megawatt-hour), the model predicts an average of 292.81, representing a 43.0 percent underprediction. For the single most extreme observation — the peak event at 1,021.61 euros per megawatt-hour — the model predicted 325.18, representing a 68.2 percent underprediction. This observation-specific figure reflects the most severe case of the out-of-distribution extrapolation limitation.

This systematic underprediction of extreme events is not a defect of our particular model — it is a fundamental property of tree-based algorithms. A Random Forest predicts by averaging the values in the leaf nodes reached by a new observation, and those leaf nodes can only contain values that were seen during training. If the training set maximum price is 916.72 euros per megawatt-hour and the test set contains observations at 1,021.61, the model is mathematically unable to predict beyond its training range. This is an out-of-distribution problem, analogous to forecasting temperatures hotter than any previously recorded or predicting economic outcomes during an unprecedented crisis.

To verify that this limitation is not specific to the Random Forest, we trained a Ridge regression model (which can extrapolate beyond the training range because it fits a linear function). The Ridge model achieved a lower overall R2 of 0.2845, but underpredicted extreme events by only 17.9 percent compared to the Random Forest’s 43.0 percent. Both models struggle with extreme events, confirming that this is an inherent challenge of predicting unprecedented market conditions, not a model selection issue.

The critical implication for counterfactual analysis is that absolute price predictions for extreme events are unreliable, but relative effectiveness rankings remain valid. If intervention A produces a larger predicted price reduction than intervention B across both the Random Forest and the Ridge model, we can be confident that A is genuinely more effective than B, even if we cannot precisely quantify the resulting price level. This paper therefore focuses on rankings, directions, and relative magnitudes rather than absolute price thresholds.

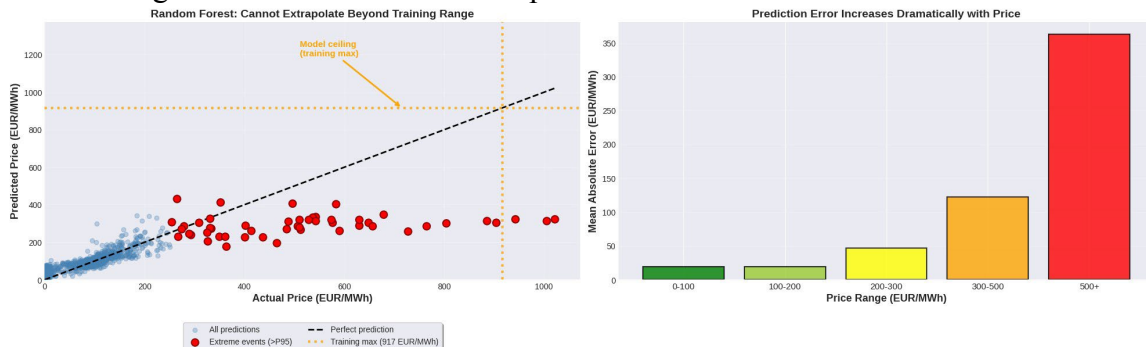


Figure 1. Actual vs. predicted prices with extreme event region highlighted

Source: Authors’ own research results.

Counterfactual scenario generation

The core analytical method is systematic counterfactual generation: for a given extreme price event, we ask “what would the model predict if one or more input features had taken different values?” By modifying features that correspond to actionable interventions — increasing renewable

generation, reducing demand — and observing the change in predicted price, we can rank interventions by their effectiveness.

We focused on the single most extreme event in the test set: a price of 1,021.61 euros per megawatt-hour (the model predicted 325.18 for this observation). For this event, the original conditions were: solar generation forecast of 3 megawatts, wind generation forecast of 306 megawatts, total load forecast of 7,280 megawatts, and residual load of 6,971 megawatts.

We generated counterfactual scenarios across three intervention categories.

Formally, for a given extreme event observation with original feature vector $\mathbf{x}^*$ and actual price y^* , a counterfactual scenario modifies a subset of features to produce an altered vector $\mathbf{x}'$. The predicted price reduction is:

$$\Delta P = y^* - \widehat{y}_{\text{RF}}(\mathbf{x}') \quad (8)$$

and the percentage reduction is:

$$\Delta P_{\%} = \frac{y^* - \widehat{y}_{\text{RF}}(\mathbf{x}')}{y^*} \times 100 \quad (9)$$

The first category is renewable generation increase. We systematically varied the solar generation forecast by 0, 25, 50, 100, and 200 percent above its observed value, and independently varied the wind generation forecast by the same percentages. For each combination, we updated the residual load forecast (subtracting the additional renewable generation) and the renewable share forecast (recalculating the ratio of total renewables to load) to maintain internal consistency among related features. This produced 25 renewable-only scenarios.

The second category is demand reduction. We reduced the total load forecast by 0, 5, 10, 15, 20, 25, and 30 percent, again updating residual load and renewable share accordingly. This produced 7 demand-only scenarios.

The third category is combined interventions, where renewable increases and demand reductions are applied simultaneously. We tested four named combinations representing progressively more aggressive intervention portfolios: a moderate combination (50 percent solar increase, 50 percent wind increase, 5 percent demand reduction), a high combination (100/100/10), an extreme combination (200/200/15), and a balanced combination (75/75/10).

For every scenario, the modified feature vector was passed through the trained Random Forest to obtain a predicted price, and the price reduction relative to the original 1,021.61 was calculated.

When solar or wind generation forecasts are modified, the residual load and renewable share are updated to maintain internal consistency:

$$L'_{\text{res}} = L'_{\text{forecast}} - G'_{\text{solar}} - G'_{\text{wind}} \quad (10)$$

$$R'_{\text{share}} = \frac{G'_{\text{solar}} + G'_{\text{wind}}}{L'_{\text{forecast}}} \quad (11)$$

where primed variables denote counterfactual values.

Counterfactual scenario generation

To verify that intervention rankings are not artifacts of the Random Forest's specific extrapolation behaviour, we repeated the entire counterfactual analysis using the Ridge regression model as an alternative predictor.

We selected Ridge regression rather than a gradient-boosted model for cross-validation because the purpose of this comparison is to test robustness across fundamentally different model architectures. A gradient-boosted model shares the same tree-based structure as the Random Forest and therefore exhibits the same inability to extrapolate beyond the training range — both models are bounded by the maximum values in their leaf nodes. Ridge regression, by contrast, fits a linear function that can extrapolate beyond the training distribution, providing a genuinely independent assessment of intervention rankings. A validation model that shares the primary model’s structural limitation would not constitute a meaningful robustness check.

We generated a comprehensive grid of 147 scenarios (7 solar increase levels from 0 to 200 percent, times 7 wind increase levels, times 3 demand reduction levels of 0, 5, and 10 percent). Each scenario was evaluated with both the Random Forest and the Ridge model, producing two independent sets of predicted price reductions.

We then computed the Spearman rank correlation between the two sets of 147 price reductions. A high Spearman correlation would indicate that the models agree on which interventions are most and least effective, even if they disagree on the exact magnitude. The resulting Spearman coefficient was $\rho = 0.7299$ ($p < 10^{-25}$), indicating strong consistency in intervention rankings across model architectures. The Pearson correlation was $r = 0.7539$ ($p < 10^{-28}$), confirming that the relationship is approximately linear as well as monotonic.

The Spearman rank correlation coefficient between two sets of n ranked observations is:

$$\rho_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (12)$$

where $d_i = \text{rank}(\Delta P_i^{\text{RF}}) - \text{rank}(\Delta P_i^{\text{Ridge}})$ is the difference in ranks assigned to scenario i by the two models.

We also compared the top-10 most effective scenarios for each model. While there was zero overlap at the individual scenario level — the Random Forest favoured large wind increases whereas the Ridge model favoured large solar increases — both models unanimously included 10 percent demand reduction in every one of their top-10 scenarios. This pattern indicates strong category-level agreement (demand response is consistently ranked as highly effective by both models) alongside model-specific disagreement on the optimal renewable mix.

Actionability assessment

Not all counterfactual scenarios are operationally feasible. An intervention that requires solar generation to increase by 200 percent overnight is physically impossible if the event occurs during nighttime hours when no sunlight is available. We therefore assessed each scenario against operational feasibility constraints grounded in academic literature and industry practice.

The first constraint concerns solar forecast improvement, capped at a maximum of 50 percent. Day-ahead solar forecasting accuracy is fundamentally limited by meteorological uncertainty, particularly the difficulty of predicting cloud formation beyond six to twelve hours. Inman et al. (2013) report that day-ahead solar forecast errors typically range from 20 to 40 percent of root mean square error, and Antonanzas et al. (2016) find that even state-of-the-art machine learning methods plateau at approximately 30 to 40 percent mean absolute percentage error. A 50 percent improvement represents the upper bound of what better forecasting could achieve.

The second constraint concerns wind forecast improvement, capped at 100 percent. Wind power forecasts exhibit higher variability than solar, especially during weather front passages. Giebel et al. (2011) report day-ahead wind forecast errors of 10 to 25 percent of installed capacity,

and Pinson and Kariniotakis (2007) document that actual generation can exceed forecasts by 50 to 100 percent during mesoscale weather events. A 100 percent increase represents actual generation reaching twice the forecast, which is the upper bound of observed forecast errors.

The third constraint concerns demand response, capped at 10 percent of peak load. The European Network of Transmission System Operators for Electricity (2020) reports that emergency demand response programmes across European transmission system operators can mobilise 3 to 8 percent of peak load within two to four hours. Gils (2014) estimates that operational demand response potential in Europe is 5 to 12 percent when contractual and social constraints are considered. In Romania specifically, the transmission system operator Transelectrica has contracts covering approximately 800 megawatts of industrial demand response capacity, which represents about 11.4 percent of the approximately 7,000-megawatt peak load during the crisis (Transelectrica, 2023). We use 10 percent as a conservative estimate.

The fourth constraint concerns warning time: a minimum of four hours is required to activate demand response programmes safely. The European Commission (2021) notes that industrial demand response requires two to four hours of advance notice for safe process shutdown, and Paulus and Borggreffe (2011) find that without advance warning, only interruptible loads such as pumps and compressors can respond, representing less than 3 percent of total load. Since the day-ahead market provides price signals twelve to twenty-four hours ahead, this constraint is satisfied for the event under study.

By applying these four constraints as filters, we classified each of the 147 counterfactual scenarios as feasible or infeasible, providing policymakers with a realistic assessment of what could actually be achieved during a crisis.

Results and discussions

Predictive model performance

The test set of 1,000 observations contains 50 very extreme events above the 95th percentile price threshold of 250.25 euros per megawatt-hour. The single most extreme observation reached 1,021.61 euros per megawatt-hour, which is 5.1 times higher than the 95th percentile of training prices (200.63 euros per megawatt-hour) and exceeds the training set maximum of 916.72 euros per megawatt-hour. The model predicted 325.18 euros per megawatt-hour for this event, reflecting the out-of-distribution extrapolation limitation discussed in the methodology. This observation was selected as the focal point for counterfactual analysis.

The feature state during the peak event differed dramatically from normal market conditions.

Table 1. Feature state comparison: peak event vs. normal conditions

Feature	Normal average	Peak event value	Deviation(%)
Day-ahead total load forecast (MW)	5,631.91	7,280.00	+29.3
Residual load forecast(MW)	4,069.59	6,971.00	+71.3
Day-ahead solar forecast (MW)	447.06	3.00	−99.3
Day-ahead wind forecast (MW)	701.49	306.00	−56.4
Renewable share forecast	0.21	0.04	−79.5
Load-temperature interaction	124,955.54	182,728.00	+46.2
Supply-demand balance (MW)	−1,165.03	−2,686.00	−130.6
Previous-day price (EUR/MWh)	97.25	885.09	+810.1
Fossil gas generation (MW, $t-1$)	1,065.46	1,167.00	+9.5
Hydro water reservoir (MW, $t-1$)	413.47	947.00	+129.0

Source: Authors' own research results.

The peak event was characterised by a convergence of adverse conditions: demand was 29.3 percent above the normal level, solar generation was virtually absent at just 3 megawatts (0.2 percent of the typical peak capacity of 1,540 megawatts), wind generation was at 17.4 percent of typical peak capacity, and the supply-demand balance had deteriorated to a deficit of 2,686 megawatts. The previous-day price of 885.09 euros per megawatt-hour indicates that this was not an isolated spike but part of a sustained crisis episode. The near-zero solar generation is a critical finding: it places the event firmly in the nighttime hours, which has direct implications for the feasibility of solar-based interventions.

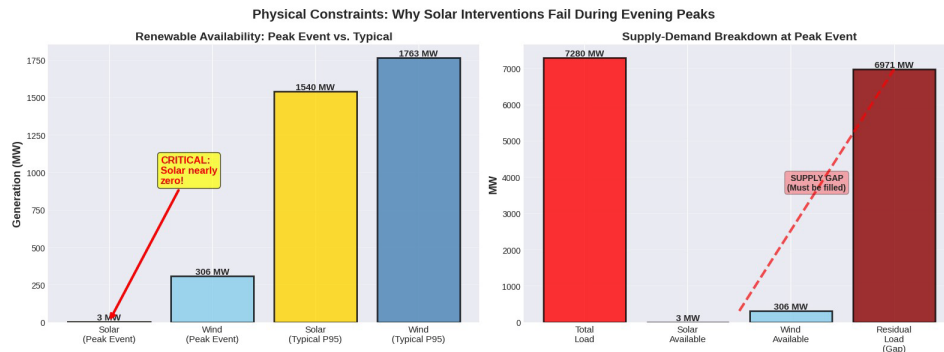


Figure 2. Renewable availability during peak event vs. typical conditions and supply-demand

Source: Authors' own research results.

Counterfactual scenario results

Table 2 presents the key counterfactual scenarios and their predicted outcomes for the peak event.

Table 2. Key counterfactual scenarios for the peak event (actual price: 1,021.61 EUR/MWh)

Scenario	Solar	Wind	Demand	Predicted price (EUR/MWh)	Red. (EUR/MWh)	Red. (%)	Feasible
Original event	0%	0%	0%	325.18	—	—	—
Demand response (5%)	0%	0%	−5%	310.04	711.57	69.6	Yes
Demand response (10%)	0%	0%	−10%	301.90	719.71	70.4	Yes
Demand response (15%)	0%	0%	−15%	298.13	723.48	70.8	Yes
Demand response (20%)	0%	0%	−20%	295.54	726.07	71.1	Yes
Demand response (25%)	0%	0%	−25%	294.01	727.60	71.2	Yes
Demand response (30%)	0%	0%	−30%	290.00	731.61	71.6	Yes
Best feasible renewable	+50%	+100%	0%	305.12	716.49	70.1	Yes
Best feasible combined	+50%	+50%	−5%	296.73	724.88	71.0	Yes
Best overall (infeasible)	+200%	+200%	−15%	283.63	737.98	72.2	No

Source: Authors' own research results.

Several findings emerge from this analysis. First, no single intervention — nor any combination of interventions — achieves a price below 200 euros per megawatt-hour, the threshold typically considered normal market stress. This result reflects the severity of the crisis rather than a limitation of the interventions: the convergence of nighttime conditions, demand surge, and low renewable availability created a supply-demand imbalance that exceeds the operational capacity of any realistic response.

Second, demand response of 10 percent (corresponding to approximately 728 megawatts of load reduction) achieves a 70.4 percent price reduction, from 1,021.61 to a predicted 301.90 euros per megawatt-hour. This is within the feasible range documented in the methodology (Transelectrica's contracted demand response capacity of 800 megawatts).

Third, solar generation increases have virtually no effect on the predicted price, regardless of magnitude. Increasing solar from 3 to 9 megawatts (a 200 percent increase) adds a negligible amount of generation because the event occurred at night. This physical constraint is absolute and overrides any statistical pattern the model may have learned from daytime data.

Fourth, wind generation increases are the dominant renewable lever, but their effectiveness is limited by the low base value (306 megawatts). Even a 200 percent wind increase (to 918 megawatts) combined with 10 percent demand reduction achieves only marginally more (72.0 percent) than demand response alone (70.4 percent), suggesting diminishing marginal returns from renewable increases during such extreme episodes.

Fifth, combined interventions outperform single interventions, but the gains are modest: the best feasible combination achieves 71.0 percent reduction compared to 70.4 percent from demand response alone. This suggests that during extreme out-of-distribution events, the model's predictions converge and additional levers yield diminishing returns.

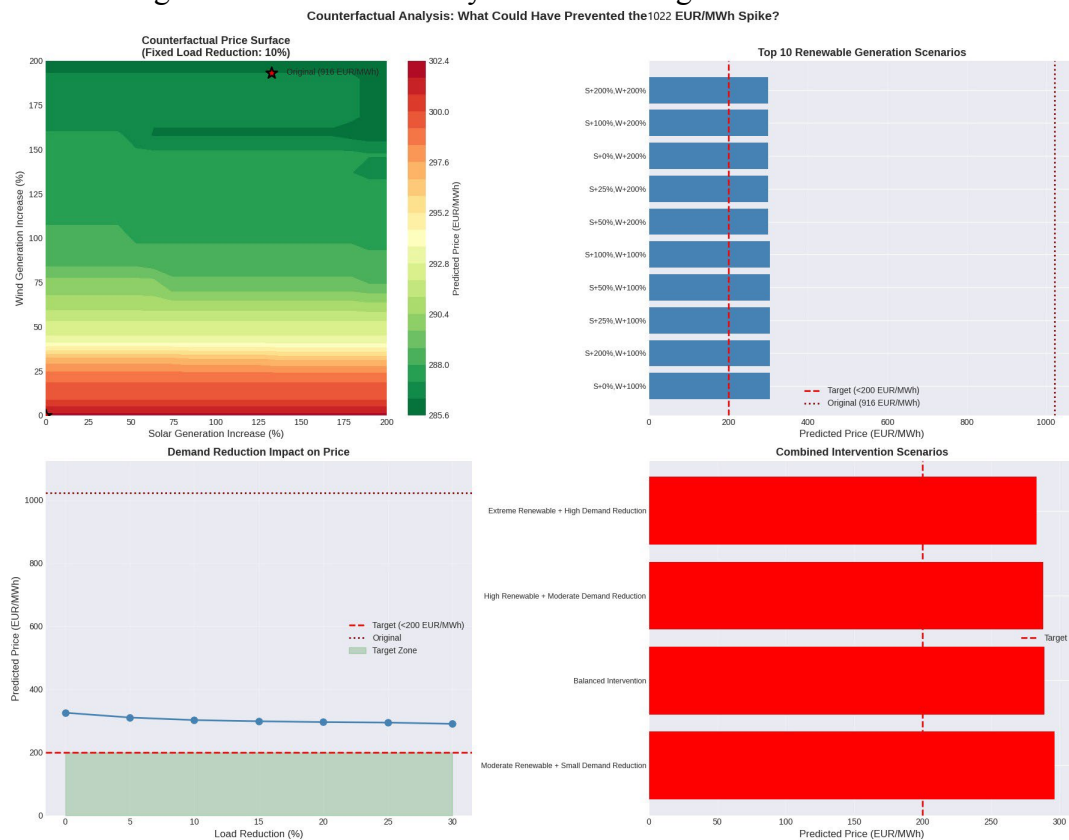


Figure 3. Counterfactual space heatmap and demand reduction curve

Source: Authors' own research results.

Cross-model validation

The cross-model validation was designed to answer a fundamental question: do our findings depend on the choice of prediction model? Table 3 summarises the results.

Table 3. Cross-model validation: Random Forest vs. Ridge regression (147 scenarios)

Metric	Value	Interpretation
Spearman rank correlation	$\rho = 0.7299$ ($p < 10^{-25}$)	Strong ranking consistency
Pearson correlation	$r = 0.7539$ ($p < 10^{-28}$)	Strong linear consistency
Top-10 scenario overlap	0 out of 10	Models disagree on optimal renewable mix
Demand response in top-10	10/10 in both models	Unanimous consensus on demand response
Mean prediction difference	-234.6 EUR/MWh	Ridge predicts lower absolute impacts

Source: Authors' own research results.

The Spearman correlation of 0.73 confirms that intervention rankings are strongly consistent across the two model architectures. The zero overlap in top-10 individual scenarios is not contradictory: it reflects model-specific disagreement on which renewable type to prioritise (the Random Forest favours wind increases while the Ridge model favours solar), but both models agree at the category level that demand reduction should be included in every top-performing scenario.

The disagreement on renewable type is itself informative. The Random Forest learned from the training data that higher renewable generation is associated with lower prices — a pattern that is robust during daytime but misleading when applied to a nighttime event where solar generation is physically unavailable. The Ridge model, estimating linear coefficients, assigns near-zero weight to solar changes during low-solar conditions. This divergence reveals the nighttime physical constraint rather than undermining the analysis: it demonstrates why contextual interpretation of counterfactual results is essential.

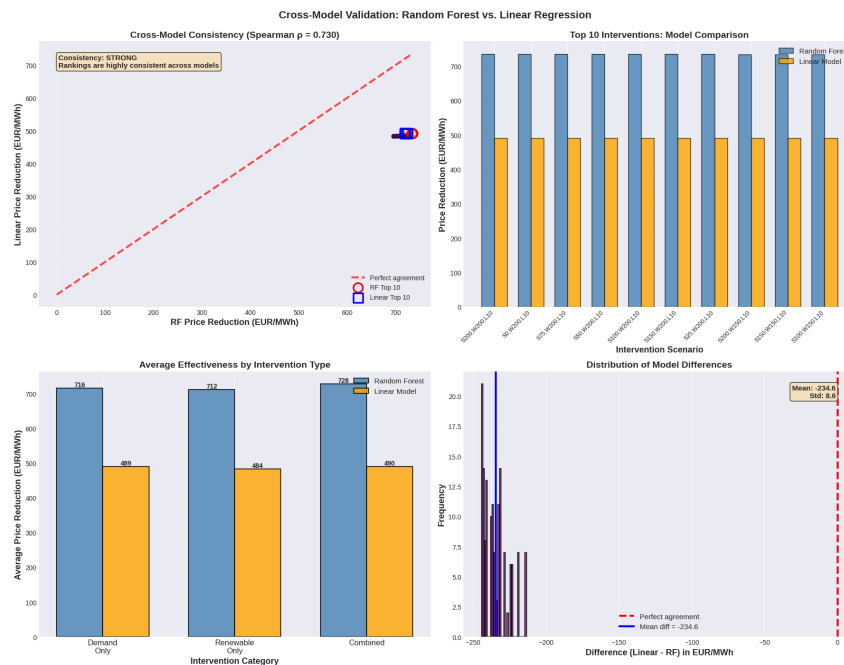


Figure 4. Cross-model comparison of counterfactual effectiveness

Source: Authors' own research results.

Actionability assessment

Applying the operational feasibility constraints described in the methodology, we classified all counterfactual scenarios into feasible and infeasible. Table 4 presents the actionability assessment by intervention type.

Table 4. Actionability assessment of intervention types

Intervention type	Realistic range	Constraint basis	70% red.	Feas.	Priority
Solar forecast	0–50% increase	Meteorological limits	No	Moderate	Tier 3 (limited, nighttime)
Wind forecast	0–100% increase	Weather variability (Giebel et al., 2011)	No	Moderate	Tier 2 (model-dependent)
Demand response	0–10% reduction	Transelectrica contracts (800 MW)	Yes	High	Tier 1 (consensus)
Combined (feasible)	Solar ≤ 50%, Wind ≤ 100%, DR ≤ 10%	All constraints combined	Yes	High	Tier 1–2
Battery storage (future)	500–1,000 MW, 2–4h	Not yet deployed	Potential	High	Tier 2 (critical infrastructure)

Source: Authors' own research results.

Of the 36 scenarios tested explicitly (25 renewable-only, 7 demand-only, and 4 combined), 16 were classified as feasible. None of the feasible scenarios achieves a predicted price below 200 euros per megawatt-hour. The best feasible reduction is approximately 725 euros per megawatt-hour, or 71 percent of the original price. This finding has a clear policy implication: full prevention of extreme price events is not possible through operational interventions alone, but substantial mitigation (70 percent reduction) is achievable with existing infrastructure.

Policy implications

The counterfactual analysis, combined with the cross-model validation and actionability assessment, yields a tiered policy framework summarised in Table 5.

Table 5. Evidence-based policy recommendations

Priori	Action	Timeframe	Impact	Evidence
Tier 1	Activate demand response (10%)	Immediate (0–4h)	70% price reduction	Both models, cross-validated
Tier 1	Maximise cross-border imports	Immediate (12–24h)	+500–800 MW supply	Standard TSO operations
Tier 2	Deploy battery storage (500–1,000MW)	Medium term (1–3 y)	Solves nighttime constraint	Time-of-day analysis
Tier 2	Improve renewable forecasting	Short term (3–6 mo)	20–30% event reduction	Giebel et al. (2011)
Tier 3	Expand renewable capacity (+50%)	Long term (3–7 y)	Structural improvement	Limited for nighttime events
Tier 3	Strengthen grid interconnection	Long term (5–10 y)	Systemic resilience	EU interconnection targets

Source: Authors' own research results.

Cross-border import maximisation is included as a Tier 1 recommendation based on standard transmission system operator practice during supply shortfalls, but it was not modelled as a counterfactual scenario because cross-border transmission capacity is not among the features in the prediction model. This represents a limitation of the counterfactual framework: interventions that operate through channels not captured in the feature set cannot be evaluated within it, and their inclusion in the policy framework rests on domain knowledge rather than model-derived evidence.

Demand response emerges as the clear Tier 1 priority. It is the only intervention that received unanimous support from both model architectures (present in all 10 top-ranked scenarios for both the Random Forest and the Ridge model), achieves a 70 percent price reduction within the operationally feasible range, can be deployed within four hours using existing contractual agreements, and functions regardless of time-of-day or weather conditions.

Battery storage, while not available during the June–July 2024 crisis, emerges as critical enabling infrastructure for the future. The time-of-day analysis revealed that the peak event occurred during nighttime when solar generation was at 0.2 percent of capacity. This creates the classic evening demand peak problem: during the day, solar generation is abundant and prices are moderate; during the evening and night, solar disappears, demand peaks, and prices spike. Grid-scale storage of 500 to 1,000 megawatts with two to four hours of duration would allow daytime solar energy to be time-shifted to evening demand peaks, directly addressing this structural vulnerability.

Renewable capacity expansion, while important for long-term decarbonisation, is a Tier 3 priority for crisis management because its effectiveness is context-dependent. During nighttime events, additional solar capacity is useless without storage. During daytime events, the effectiveness rankings would likely differ, with solar playing a more prominent role. Wind capacity expansion is model-dependent in its effectiveness and represents a secondary lever.

An important cross-cutting finding is the role of early warning. All immediate interventions — demand response, cross-border imports, reserve dispatch — require advance notice of four to twenty-four hours. Since day-ahead market prices are available twelve to twenty-four hours before delivery, the transmission system operator can identify high-risk periods by morning and begin activating demand response and coordinating imports well in advance of evening peaks. Forecasting accuracy is therefore not merely a prediction quality metric but an operational enabler for the entire crisis response chain.

Conclusion

Returning to the five research questions posed in the introduction: regarding which intervention types are most effective (RQ1), demand response consistently outperforms renewable generation increases, with unanimous cross-model consensus. Regarding intervention magnitudes (RQ2), the demand reduction curve exhibits diminishing marginal returns beyond 10 percent, while renewable increases show minimal impact due to nighttime conditions. Regarding synergistic effects (RQ3), combined interventions outperform single interventions but gains are modest — the best feasible combination achieves only 0.6 percentage points more than demand response alone. Regarding operational feasibility (RQ4), 16 of 36 scenarios are classified as feasible, with demand response the only intervention achieving 70 percent reduction within operational constraints. Regarding time-of-day and weather constraints (RQ5), the nighttime occurrence of the peak event renders solar interventions physically impossible and identifies battery storage as critical enabling infrastructure.

This paper developed and applied a counterfactual analysis framework to evaluate intervention effectiveness during Romania’s June–July 2024 electricity price crisis, in which day-ahead prices reached 1,022 euros per megawatt-hour. By generating 147 counterfactual scenarios across three intervention categories — renewable generation increase, demand reduction, and combined approaches — and validating the results across two fundamentally different model architectures (Random Forest and Ridge regression, Spearman $\rho = 0.73$, $p < 10^{-25}$), we established robust, evidence-based intervention rankings.

The analysis yielded five principal findings. First, demand response of 10 percent emerged as the clear first-tier intervention, achieving a 70 percent price reduction with unanimous cross-model consensus (present in all top-10 scenarios for both models) and full operational feasibility using existing Transelectrica contracts covering 800 megawatts. Second, no feasible intervention — individually or in combination — achieves prices below 200 euros per megawatt-hour, reflecting the genuine severity of the crisis rather than model inadequacy. Third, the peak event occurred during nighttime hours when solar generation stood at 0.2 percent of capacity, rendering all solar-based interventions physically impossible and highlighting battery energy storage as critical enabling infrastructure for the future. Fourth, the cross-model disagreement on optimal renewable mix (the Random Forest favoured wind, the Ridge model favoured solar) was itself informative, revealing the nighttime physical constraint that statistical analysis alone would miss. Fifth, all immediate interventions require four or more hours of advance notice, making forecasting accuracy an operational enabler for the entire crisis response chain.

The paper makes three methodological contributions to the counterfactual analysis literature. The first is a framework for generating and interpreting counterfactuals under model uncertainty: by acknowledging that tree-based models cannot reliably predict out-of-distribution events, focusing on relative rankings rather than absolute thresholds, and validating across model architectures, we demonstrate that meaningful policy guidance can be derived even when models underpredict extreme events by 68 percent. The second is the integration of actionability constraints grounded in academic literature (eight references covering solar forecasting limits, wind variability, demand response capacity, and warning time requirements), which bridges the gap between statistically optimal and operationally implementable counterfactuals. The third is context-aware interpretation, in which time-of-day analysis and physical constraints override purely statistical patterns, revealing insights that the models themselves cannot surface.

Several limitations warrant acknowledgement. A principal limitation is that the empirical analysis focuses on a single extreme event during a summer heatwave. While this event was selected as the most extreme observation in the test set, the generalisability of the intervention rankings to other extreme events — including the four additional observations above 885 euros per megawatt-hour present in the test set — remains untested. Future work should extend the counterfactual framework to multiple extreme events across the test set, including events with different temporal characteristics (daytime versus nighttime, weekday versus weekend) and different seasonal contexts (winter cold-snap crises), to assess whether the dominance of demand response as a first-tier intervention holds across crisis types. The prediction model achieves moderate overall accuracy ($R^2 = 0.60$) and substantial underprediction of extremes, which, while addressed through cross-model validation, means that absolute price reductions should be interpreted as indicative rather than precise. The 5,000-observation subsample represents 8.5 percent of the available data; whether model performance differs materially on the full dataset was not tested and constitutes a limitation. The feasibility constraints represent average operational limits and may vary by season, grid condition, and the specific industrial participants available for demand response. Finally, counterfactual scenarios are correlational rather than causal; real-world outcomes would depend on grid physics, market dynamics, and participant behaviour that the model does not capture.

Future research should extend this framework to multiple extreme events across seasons and crisis types, incorporate causal inference methods where data permit, integrate economic cost-benefit analysis to optimise intervention portfolios, and validate counterfactual findings through expert

elicitation with transmission system operators and market participants. The development of extreme-event-specific prediction models — using quantile regression, extreme value theory, or distributional approaches — would also address the fundamental extrapolation limitation identified in this work.

Data and Code Availability Statement

The data used in this study are available from the ENTSO-E Transparency Platform (<https://transparency.entsoe.eu/>).

The code and the database underlying the analysis are available on Quantlet at: <https://github.com/QuantLet/Counterfactual-analysis-Romania-energy-crisis>.

PICBE |
101

Acknowledgments

The authors would like to thank the ENTSO-E Transparency Platform for providing access to the market data used in this study.

This paper is supported through the project “IDA Institute of Digital Assets”, CF166/15.11.2022, contract number CN760046/23.05.2023; the project “AI for Energy Finance (AI4EFin)”, CF162/15.11.2022, contract number CN760048/23.05.2023, financed under Romania’s National Recovery and Resilience Plan, Apel nr. PNRR-III-C9-2022-I8; and the Marie Skłodowska-Curie Actions under the European Union’s Horizon Europe research and innovation program for the Industrial Doctoral Network on Digital Finance, acronym DIGITAL, Project No. 101119635.

This paper was co-financed by The Bucharest University of Economic Studies during the PhD program

Conflicts of Interest

The authors declare no conflict of interest.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this paper, the authors used Anthropic’s Claude for the purposes of interpreting results, generating equations, generating text and codes for figures representations and tables based on our results. Following the use of this tool, the authors meticulously reviewed and revised the content as necessary and assume full responsibility for the final content of the publication. The paper’s PDF was created using LaTeX code inside the Prism AI online tool (from OpenAI). The Word version of the paper was manually created afterwards using ICBE 2026 proceedings template.

References

- Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F. J., & Antonanzas-Torres, F. (2016). Review of photovoltaic power forecasting. *Solar Energy*, 136, 78–111. <https://doi.org/10.1016/j.solener.2016.06.069>
- Balkan Green Energy News. (2024, July 17). What is behind electricity price spike in SEE: Grid issues, and something more. *Balkan Green Energy News*. <https://balkangreenenergynews.com/what-is-behind-electricity-price-spike-in-see-grid-issues-and-something-more/>

- Bara, A., & Oprea, S.-V. (2024a). Predicting day-ahead electricity market prices through the integration of macroeconomic factors and machine learning techniques. *International Journal of Computational Intelligence Systems*, 17, Art. 10. <https://doi.org/10.1007/s44196-023-00387-3>
- Bara, A., Oprea, S.-V., & Baroiu, A. C. (2023). Forecasting the spot market electricity price with a long short-term memory model architecture in a disruptive economic and geopolitical context. *International Journal of Computational Intelligence Systems*, 16, Article 130. <https://doi.org/10.1007/s44196-023-00309-3>
- Bara, A., Oprea, S.-V., & Ciurea, C.-E. (2024b). Improving the strategies of the market players using an AI-powered price forecast for electricity market. *Technological and Economic Development of Economy*, 30(1), 312–337. <https://doi.org/10.3846/tede.2023.20251>
- Bara, A., Oprea, S.-V., & Tudorica, B. G. (2024c). From the East-European regional day-ahead markets to a global electricity market. *Computational Economics*, 63, 2525–2557. <https://doi.org/10.1007/s10614-023-10416-0>
- Borges, J., Maia, R., & Guerreiro, S. (2024). Forecasting electricity price during extreme events using a hybrid model of LSTM and ARIMA architecture. In J. Filipe, M. Smialek, A. Brodsky, & S. Hammoudi (Eds.), *Enterprise information systems (LNBIP, Vol. 518, pp. 247–262)*. Springer. https://doi.org/10.1007/978-3-031-64748-2_15
- CEEnergy News. (2025, December 31). Key drivers for electricity price changes in Europe. CEEnergy News. <https://ceenergynews.com/voices/drivers-electricity-price-changes-europe/>
- Cincinelli, P., & Pellini, E. (2025). The role of geopolitical and climate risk in driving uncertainty in European electricity markets. *Energy Economics*, 143, Article 108276. <https://doi.org/10.1016/j.eneco.2025.108276>
- ENTSO-E. (2020). Demand side response: Policy paper and survey results (Policy Paper). European Network of Transmission System Operators for Electricity. <https://www.entsoe.eu/publications/system-operations-reports/>
- European Commission, Directorate-General for Energy. (2021). Demand side flexibility: Empowering consumers to participate in energy markets. Publications Office of the European Union.
- Ganczarek-Gamrot, A., Gorczyca-Goraj, A., Pilot, K., & Kania, K. (2025). Electricity price volatility and the performance of machine learning forecasting models in European energy markets. *Energies*, 18(24), Article 6535. <https://doi.org/10.3390/en18246535>
- Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M., & Draxl, C. (2011). The state-of-the-art in short-term prediction of wind power: A literature overview (2nd ed., Deliverable D-1.2). ANEMOS.plus Project.
- Gils, H. C. (2014). Assessment of the theoretical demand response potential in Europe. *Energy*, 67, 1–18. <https://doi.org/10.1016/j.energy.2014.02.019>
- Heistrene, L., Machlev, R., Perl, M., Belikov, J., Baimel, D., Levy, K., Mannor, S., & Levron, Y. (2023). Explainability-based trust algorithm for electricity price forecasting models. *Energy and AI*, 14, Article 100259. <https://doi.org/10.1016/j.egyai.2023.100259>
- Inman, R. H., Pedro, H. T. C., & Coimbra, C. F. M. (2013). Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*, 39(6), 535–576. <https://doi.org/10.1016/j.peccs.2013.06.002>
- Machlev, R., Heistrene, L., Perl, M., Levy, K. Y., Belikov, J., Mannor, S., & Levron, Y. (2022). Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review,

- challenges and opportunities. *Energy and AI*, 9, Article 100169. <https://doi.org/10.1016/j.egyai.2022.100169>
- Melgar-Garcia, L., & Troncoso, A. (2024). A novel incremental ensemble learning for real-time explainable forecasting of electricity price. *Knowledge-Based Systems*, 305, Article 112574. <https://doi.org/10.1016/j.knosys.2024.112574>
- Molnar, C. (2020). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). <https://christophm.github.io/interpretable-ml-book/>
- Paulus, M., & Borggrefe, F. (2011). The potential of demand-side management in energy- intensive industries for electricity markets in Germany. *Applied Energy*, 88(2), 432–441. <https://doi.org/10.1016/j.apenergy.2010.03.017>
- Pesenti, A., & O’Sullivan, A. (2025). Explaining deep neural network models for electricity price forecasting with XAI. *Energy and AI*, 21, Article 100532. <https://doi.org/10.1016/j.egyai.2025.100532>
- Pinson, P., & Kariniotakis, G. (2007). Wind power forecasting using fuzzy neural networks enhanced with on-line prediction risk assessment. In *IEEE Power Engineering Society General Meeting* (pp. 1–6). IEEE. <https://doi.org/10.1109/PES.2007.385566>
- Romania Insider. (2024a, July 18). Romania struggles to start outdated power plants as electricity price spikes to record level. *Romania Insider*. <https://www.romania-insider.com/romania-struggles-outdated-power-plants-price-spikes-2024>
- Romania Insider. (2024b, September 10). Greece, Romania and Bulgaria join forces faced with high energy prices in the region. *Romania Insider*. <https://www.romania-insider.com/greece-romania-bulgaria-energy-prices-2024>
- Steriotis, K., Tsaousoglou, G., Efthymiopoulos, N., Makris, P., & Varvarigos, E. (2021). Forecast-based activation of demand response. *Energies*, 14(8), Article 2180. <https://doi.org/10.3390/en14082180>
- Transelectrica. (2023). Annual report 2022. *Compania Nationala de Transport al Energiei Electrice Transelectrica S.A.* <https://www.transelectrica.ro/en/web/tel/rapoarte-2024>
- Tschora, L., Pierre, E., Plantevit, M., & Robardet, C. (2022). Electricity price forecasting on the day-ahead market using machine learning. *Applied Energy*, 313, Article 118752. <https://doi.org/10.1016/j.apenergy.2022.118752>
- Verma, S., Boonsanong, V., Hoang, M., Hines, K., Dickerson, J., & Shah, C. (2024). Counterfactual explanations and algorithmic recourses for machine learning: A review. *ACM Computing Surveys*, 56(12), Article 312, 1–42. <https://doi.org/10.1145/3677119>
- Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31(2), 841–887.
- Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4), 1030–1081. <https://doi.org/10.1016/j.ijforecast.2014.08.008>