

ORIGINAL RESEARCH

Open access

Advancing Electricity Price Forecasting with Conditional Generative Adversarial Networks: A Probabilistic Simulation Framework

Lina Hassan^{1*}, Omar Khalaf², Reem Jaber¹

Abstract

A new method for generating time series simulations of electricity price scenarios is introduced. The “Time Series Simulation Conditional Generative Adversarial Network” (TSS-CGAN) produces short-term scenarios of electricity prices. Specifically, the network can output a 24-dimensional vector that reflects the typical dynamics of electricity markets. This approach can serve as an alternative to conventional financial-mathematical techniques, such as statistical factor models, for modeling the price distribution around a given forecast. The dataset spans 3 years from 2020 to 2023. The empirical analysis focuses on the EPEX SPOT market in Europe. An electricity price scenario consists of the prices for the hourly contracts from a day-ahead auction on the EPEX SPOT power exchange. The model takes multivariate time series as input features, including point forecasts of electricity prices along with fundamental data on generation and load profiles. The TSS-CGAN architecture integrates the principles of Conditional Generative Adversarial Networks with 1D Convolutional Neural Networks and Bidirectional Long Short-Term Memory units. Model performance is assessed using both qualitative and quantitative measures. For evaluation purposes, 10,000 simulations are generated for a test period. Qualitative assessment examines whether the model correctly captures and reproduces key electricity market-specific patterns. The quantitative evaluation involves comparing standard error metrics against benchmark models such as DeepAR, Prophet, and Temporal Fusion Transformer, analyzing quantile ranges and error distributions, and conducting a sensitivity analysis. The findings demonstrate that the TSS-CGAN surpasses benchmark models like DeepAR by lowering the continuous ranked probability score by 50% while effectively accounting for market-specific factors, such as the generation of intermittent renewable energies, and responding appropriately to variations in the relevant variables.

Keywords Generative adversarial networks, Time series simulation, Probabilistic modeling, Day-ahead electricity prices, 1D convolutions, Bidirectional long short-term memory

*Correspondence:

Lina Hassan
lina.hassan@ju.edu.jo

¹ Department of Smart Electrical Networks, Faculty of Engineering, University of Jordan, Amman, Jordan

² Department of Renewable Energy Systems, Faculty of Engineering, Jordan University of Science and Technology, Irbid, Jordan

Introduction

For over 30 years, electricity has been traded as an exchange-listed commodity in most countries worldwide. It is traded across various time horizons (forward, day-ahead, intraday), with the day-ahead market generally serving as the primary reference. Prominent examples include the European day-ahead markets on EPEX Spot, NORDPOOL in the Nordic region, and PJM or Nodal Exchange in the United States. Each electricity market exhibits distinct features. For instance, the German market is heavily influenced by variable renewable output from wind and solar, the French market relies

predominantly on large-scale nuclear facilities, and NORDPOOL is characterized by dominant hydropower generation. The makeup of the generation mix plays a critical role in shaping price behavior across different markets. The objective of this study is to create a generative model that produces realistic day-ahead electricity price scenarios, offering a potential replacement for traditional probabilistic models and their typically intricate calibration procedures [1]. A fully data-driven approach allows analysts to model diverse markets effectively without requiring extensive statistical investigation. In this paper, the German day-ahead market serves as the reference case study.

Machine learning methods significantly broaden the scope for capturing intricate relationships. Years of research have yielded numerous model architectures capable of addressing a diverse array of challenges. The continued advancement of these technologies is expected to play an increasingly important role in guiding the energy sector toward data-driven decision-making that addresses market complexity with greater accuracy and reliability.

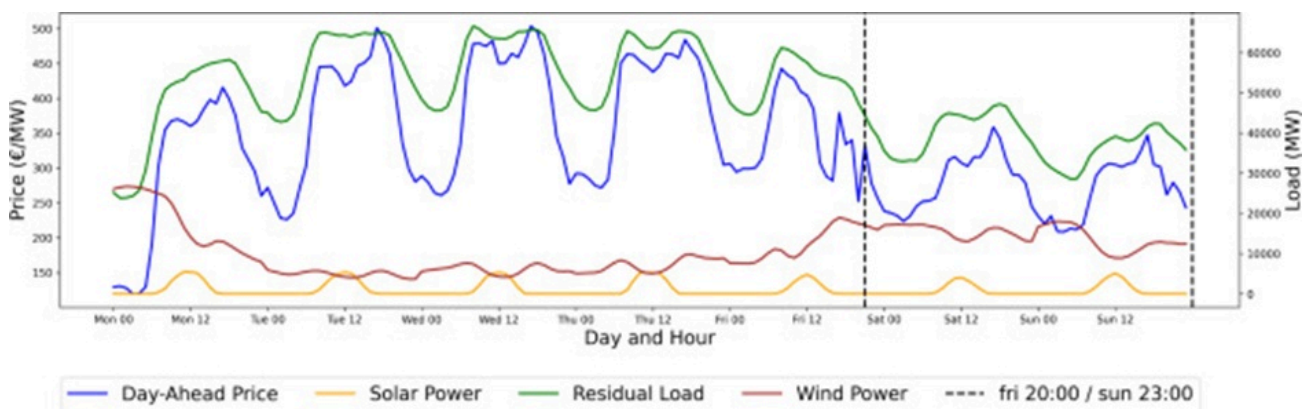


Figure 1. Characteristic weekly curve.

Electricity pricing on the day-ahead market

The day-ahead electricity price serves as a key indicator within the energy market. It represents the price established for the supply of electrical energy to be delivered on the following day. This subsection examines the formation process of the day-ahead electricity price, the factors that affect it, and the underlying market mechanisms. It also highlights the major exchanges and participants involved in price determination.

The day-ahead electricity price is established on the spot market operated by EPEX SPOT (European Power Exchange), where energy is traded across 13 European countries. Prices are set via a daily auction that matches supply and demand. This auction occurs every day at 12:00 noon (CET/CEST) and determines prices for each hour of the next day. These prices are fixed ex ante for each hour ahead, meaning the price at 00:00 covers the interval from 00:00 to 01:00. Market participants submit their bids by 12:00 (CET/CEST), indicating their readiness to purchase or sell specific quantities at particular times. The exchange constructs a demand curve from buy bids and a supply curve from sell bids, both defined for each delivery period (i.e., each of the 24 hours of the day). The point at which these curves intersect yields the market-clearing price (MCP), which is uniform across buyers and sellers. Through this matching process, the power exchange creates legally enforceable contracts for the purchase or sale of a specified volume of electricity in a designated delivery zone at the determined price. Pricing follows the merit order principle. This principle ensures the lowest possible cost to satisfy demand by prioritizing generation with the lowest marginal costs (or the lowest willingness to sell). This establishes a ranked order of generation sources, known as the “merit order.”

In Germany, the day-ahead electricity price is influenced by various elements. Weather patterns play a major role in renewable energy output: favorable conditions, such as strong winds or abundant sunshine, result in lower prices. In

contrast, unfavorable weather increases reliance on more expensive fossil fuels, thereby raising prices. Furthermore, costs of fossil fuels—such as coal, natural gas, and oil—directly impact generation expenses and, consequently, the day-ahead price.

A distinctive aspect of the German spot electricity market is the substantial share of renewable energy in the generation mix (56% in 2023 [2]). This effect is evident in electricity price patterns, as illustrated in **Figure 1**.

Simulating scenarios of uncertain electricity prices is valuable to multiple stakeholders in the energy sector. Energy suppliers employ simulation models to enhance power plant operations through stochastic optimization [3]. Risk managers use simulations to assess short-term risks, while retailers routinely use them to optimize their portfolios [4].

In practice, reasonably precise spot price forecasts are readily available, often acquired from various providers. For the applications mentioned—particularly with flexible portfolios (e.g., hydropower)—the distribution surrounding these point forecasts is especially relevant. Such distributions can be derived from fundamental models incorporating stochastic inputs [5] or from stochastic price models. Both approaches typically require either significant, often unrealistic, assumptions or extensive historical data combined with highly detailed market representations. Our method, grounded in generative models, reduces reliance on assumptions and in-depth historical scrutiny by learning market characteristics directly from past data, which are then applied to produce simulations.

Generative time series simulation

In recent years, Generative Deep Learning has attracted considerable interest [6]. It has found particularly widespread applications in image and text processing domains [7]. Technologies such as deep fakes and advanced voice assistants are now commonplace, drawing both enthusiasm and criticism. Generative models excel at replicating intricate data structures, including images or natural language sentences. The outputs they produce closely resemble the training data yet remain novel and varied. Generative modeling differs from discriminative modeling. Discriminative approaches are tied to supervised learning, in which a function is learned to map inputs to outputs using labeled datasets.

In contrast, generative techniques generally operate without labels, classifying them as a form of unsupervised learning. Generative Adversarial Networks (GANs) have become a standard method for creating synthetic data. Synthetic data are artificially generated but retain structural and statistical properties similar to those of real data. For instance, GANs can generate highly realistic images that are difficult for humans to distinguish from authentic ones. Images are commonly represented as 4D tensors (with a single image as a 3D tensor) [8]. More recently, GANs have proven effective in synthesizing lower-dimensional tensors as well. Time series modeling, in particular, has drawn growing attention. Applications range from sensor measurements [9] and stock market indices [10] to solar energy system outputs [11], demonstrating the growing popularity of GANs for time series generation.

The contributions of this paper can be summarized as follows:

- We introduce a CGAN-based architecture designed to simulate realistic electricity price scenarios, integrating forecasts of energy-market-specific features.
- We propose an evaluation framework tailored to the specific characteristics of energy markets, enabling quantification of the risk associated with a given price forecast through a generated distribution.
- We analyze the strengths and limitations of the proposed architecture and benchmark its performance against well-established probabilistic models, including DeepAR, Prophet, and the Temporal Fusion Transformer (TFT).

Related Work

Generative time series simulation

Generative Adversarial Networks (GANs), originally proposed by Ian Goodfellow and colleagues in 2014, represent a distinctive neural network framework comprising two competing networks: a generator and a discriminator. The generator produces synthetic data from random noise vectors, while the discriminator distinguishes real from generated samples. The generator's objective is to deceive the discriminator into misclassifying synthetic data as authentic. Through this adversarial training process, GANs learn to produce highly realistic, high-fidelity data that closely resembles the real distribution.

Late in 2014, Mehdi Mirza and Simon Osindero extended the original GAN framework by introducing a conditional variant [12]. Conditional GANs (CGANs) enable generation conditioned on additional information. The generator receives both a noise vector and a specific condition as input, allowing controlled output generation. Their seminal CGAN demonstration utilized the MNIST dataset, successfully producing images of handwritten digits corresponding to the provided class label. In this setup, the generator implicitly captures the data distribution through the discriminator's loss, which is optimized using both real and synthetic examples.

In recent years, GANs have been successfully applied to time-series generation across diverse domains. A prominent contribution came in 2019 from Jinsung Yoon and co-authors [13], who developed a method for generating multivariate time series while explicitly accounting for temporal correlations among variables [13]. That same year, Alireza Koochali and colleagues introduced a novel GAN-based approach for time series forecasting, training the model to predict a single future time step conditioned on the previous K steps [9]. Their architecture builds on CGAN principles, with the past K observations serving as the conditioning input to the generator. However, this method is limited to predicting one step and restricts conditioning to univariate sequences. When modeling electricity prices, leveraging multivariate time-series as covariates is beneficial, commonly incorporating factors such as weather-driven renewable generation [14], load profiles [15], and fossil fuel prices [16] to capture complex nonlinear dependencies.

In 2021, Gerardo Zúñiga and Gonzalo Acuña extended Koochali's framework to allow the generation of up to 5 future time steps $t + K$ (where $0 < K < 6$) [17]. Nevertheless, the approach continued to rely solely on univariate conditioning inputs. Also in 2021, Milena Vuletic and co-authors adapted Koochali's method to financial time series, introducing a custom loss function that incorporates profit/loss metrics and the Sharpe ratio [18].

Huang *et al.* [11] advanced the field by designing an architecture capable of forecasting the next time step of solar power output, conditioned on the prior K steps of the target variable and a corresponding feature matrix of covariates. Both the historical target sequence and the covariate matrix were provided as conditioning inputs, marking a meaningful improvement over earlier univariate conditioning strategies.

Within the domain of electricity price scenario generation, several relevant studies provide a foundation for further development. Zhang and Wu [19] demonstrated and validated the effectiveness of GANs for forecasting marginal prices in wholesale electricity markets. Demir *et al.* [20] introduced a Wasserstein GAN (WGAN) to augment data for day-ahead price forecasting regression tasks, thereby reducing regression errors. Liu *et al.* applied GAN-based models to day-ahead electricity price forecasting in Australia for the first time, building on the TimeGAN framework [13] and incorporating historical weather data and realized prices while employing Wasserstein distance as the loss function [21]. This work was further refined in 2023 to enhance prediction quality and training stability [22]. However, the use of Wasserstein loss introduces challenges such as gradient clipping, whose full implications remain underexplored. Moreover, the intricate structure of TimeGAN results in substantial computational demands. Consequently, the present study adopts a more conventional CGAN architecture with binary cross-entropy loss, following the approach in [11]. Prior research on day-ahead energy demand forecasting has shown that standard CGAN-based models can achieve performance comparable to that of WGAN variants [23].

The primary objective and core contribution of this paper is to propose a CGAN-based architecture that, similar to [11], accepts multivariate time series as input and generates all 24 hours of the day-ahead auction simultaneously. This is facilitated, among other elements, by incorporating highly accurate deterministic point forecasts as described in [24]. The referenced authors utilized precise point forecasts of the target variable as input to train a probabilistic model capable of outputting the full distribution—an approach that aligns directly with the use case here, where energy market participants acquire forecasts and seek to derive distributional scenarios via generative techniques. Finally, the architecture is designed to produce a univariate time series comprising multiple consecutive time steps, consistent with the multi-step generation capability introduced in [17].

Energy industry

A range of methods for simulating uncertain day-ahead electricity price scenarios has been developed and applied within the energy sector. Several of these techniques originate in financial mathematics, which appears well-suited to this domain owing to its flexibility and adaptability.

One widely adopted technique involves stochastic models, particularly stochastic differential equations (SDEs), to generate price scenarios, as demonstrated by Schwartz and Smith [25]. These models incorporate uncertainties and variations in key market parameters, allowing the creation of simulated paths for day-ahead electricity prices and thereby enabling projections of future price behavior.

Numerous extensions and alternatives to the Schwartz-Smith framework have since been suggested. For a comprehensive overview of factor models and their application to the German electricity market, refer to [1]. At the other end of the spectrum, the energy industry frequently relies on fundamental models. These are highly detailed representations that include extensive market data and, for each delivery period, solve an optimization problem to minimize total supply costs. Owing to their high computational demands, fundamental models are generally unsuitable for generating thousands of price trajectories. Such models and the associated data are typically proprietary to consulting firms, which either sell access to the models themselves or provide forecasted price scenarios based on specific assumptions.

Generative neural networks present a promising alternative to traditional price simulation methods, thanks to their ability to learn complex patterns and structures directly from data. Discriminative neural networks have already become well established in energy market applications, as evidenced by numerous studies confirming the technology's maturity, including [26] and [27]. In contrast, generative neural networks are only beginning to gain traction in the energy domain. Recent literature in this area includes review papers, such as that by Yilmaz *et al.* [28], which assessed various GAN architectures using Turkish intraday electricity prices. Hanif *et al.* [29] showed in 2023 that GANs outperformed models such as Support Vector Machines (SVM) and Extreme Gradient Boosting (XGBoost) when applied to Canadian electricity prices. Most recently, Wang *et al.* [30] proposed a GAN-based approach for stock price prediction in the context of factor investing.

Materials and Methods

In the field of time-series modeling using neural networks, certain specialized architectures have become standard and will be examined in greater detail in the subsections that follow. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are established architectures for handling time-series data. Both are widely deployed in industrial settings, including the finance and energy sectors [31]. The subsequent subsections on 1D convolutional layers and on Bidirectional Long Short-Term Memory—focus on the structure and function of 1D Convolutional Layers and Bidirectional Long Short-Term Memory units (Bi-LSTM), which belong to the broader family of recurrent networks. Basic recurrent networks are of limited interest here due to their restricted short-term memory capacity. In contrast, the more advanced Long Short-Term Memory (LSTM) cells are far better equipped to handle long-range dependencies and mitigate issues such as vanishing gradients.

1D convolutional layer

Convolutional layers, which form the core of CNNs, function differently from traditional fully connected neural networks.

In convolutional layers, each neuron connects only to a limited region of the preceding layer (known as the local receptive field), enabling the detection of localized patterns—such as edges or corners in images. As layers deepen, these low-level features are progressively combined into more abstract, higher-level representations, forming a hierarchical spatial structure. CNNs exhibit a degree of invariance to translations, distortions, and rescaling, enabling them to reliably identify patterns across varying conditions. Unlike fully connected networks, they do not depend on fixed positional arrangements of inputs. This principle extends naturally from higher-dimensional data (e.g., images) to lower-dimensional cases (e.g., time series).

1D convolutional neural networks have proven highly effective for time series processing [32]. Their strength stems from their ability to identify local patterns and capture temporal relationships in sequential data. By applying 1D convolutional filters, these networks automatically extract meaningful features by leveraging correlations between neighboring time points. Often, a 1D max pooling layer is applied after the convolutional layers, further improving performance by reducing the dimensionality of the feature maps and computational requirements. Max pooling selects the highest value within a defined window, thereby retaining the most salient features while filtering out less relevant information. This mechanism supports efficient identification of dominant patterns and trends in time series, making 1D CNNs particularly valuable in domains such as financial forecasting, climate analysis, and signal processing [33].

The incorporation of convolutional layers within GAN frameworks for time series modeling has become a standard technique, as seen in applications such as short-term load forecasting [34] and day-ahead wind power forecasting [35].

Bidirectional long short-term memory

Long Short-Term Memory (LSTM) layers [36] are widely adopted components in deep learning for time series forecasting, largely because of their effectiveness at modeling temporal dependencies. LSTMs maintain information about previous states over extended periods, making them well-suited to sequences of varying lengths. This capability supports the representation of intricate temporal dynamics and the simultaneous capture of both short-range and long-range dependencies within the data. In GAN frameworks, LSTM layers are often used for challenging predictive tasks, such as stock price forecasting [37].

Bidirectional LSTM (Bi-LSTM) layers, introduced in 2005 by Graves and Schmidhuber [38], enable the network to process sequences in both forward and backward directions, yielding a richer, more complete understanding of the underlying patterns. This bidirectional processing enhances accuracy in tasks that benefit from a full contextual view of long-term temporal relationships, including financial time series analysis and natural language processing, as demonstrated by Siami-Namini *et al.* [39]. The core idea is that a bidirectional recurrent network can detect dependencies that a unidirectional counterpart might overlook. Graves and Schmidhuber illustrated the advantages of this approach by comparing unidirectional and bidirectional versions of both standard RNNs and LSTMs on the TIMIT dataset, which consists of broadband speech recordings from 630 speakers representing eight major American English dialects, each reading ten phonetically balanced sentences [40]. The bidirectional architectures not only converged in roughly half the number of epochs but also achieved test-set accuracies up to 5% higher.

Bidirectional RNNs are also prevalent in natural language processing (NLP), where word dependencies are not strictly forward-directed. For instance, the interpretation of a word early in a sentence may rely on information provided much later. A similar reverse temporal dependency appears in electricity markets: a forecasted high-wind period in the afternoon can influence prices in the preceding early morning hours. Market operators anticipate the upcoming availability of low-cost renewable energy and therefore avoid starting up expensive, inflexible units (such as coal plants) with long minimum run times, opting instead for more flexible (albeit costlier) options, such as gas turbines. The primary strength of bidirectional LSTM networks is their capacity to learn an additional, complementary representation of the sequence,

offering a second viewpoint on the data. Siami-Namini et al. [39] compared LSTM and Bi-LSTM performance in stock market forecasting and found that Bi-LSTM models achieved, on average, 38% greater accuracy, largely due to their ability to uncover backward-looking temporal structures missed by unidirectional models.

Alongside LSTM and Bi-LSTM cells, the Gated Recurrent Unit (GRU) exists as a computationally lighter alternative to the LSTM cell [41]. A head-to-head comparison of LSTM, Bi-LSTM, and GRU was performed by Seabe et al. [42] in the context of cryptocurrency price prediction, where Bi-LSTM consistently outperformed both GRU and standard LSTM across multiple currencies. In energy consumption forecasting, Bi-LSTM cells have similarly demonstrated superior results [43]. Bi-LSTM layers are also applied to high-dimensional time series prediction problems, such as renewable energy output forecasting—for wind power in [44] and for solar photovoltaic (PV) generation in [11]. For a detailed explanation of the inner workings of 1D CNNs and Bi-LSTM layers, the reader is referred to standard references such as Chollet [8] or Géron [45].

Generative adversarial networks

The architecture of a Generative Adversarial Network (GAN) generally comprises two interacting neural networks: a generative model (generator) and a discriminative model (discriminator). Figure 2 illustrates the interaction between the generator and the discriminator.

The generator's input layer receives a random noise vector, typically drawn from a normal distribution and formatted as a 1D tensor, which it transforms—according to its learned weights—into a synthetic output sample. This input vector corresponds to a point in an n -dimensional latent space. The latent space provides a compact, low-dimensional encoding of the high-dimensional data distribution. In essence, the generator acts as a function (z) that maps a latent vector z sampled from distribution p_z to a data-like representation. The discriminator, denoted as (x), estimates the probability that a given sample x originates from the true data distribution p_{data} [46]. Thus, the discriminator serves as a binary classifier trained to distinguish real samples from those generated by the generator.

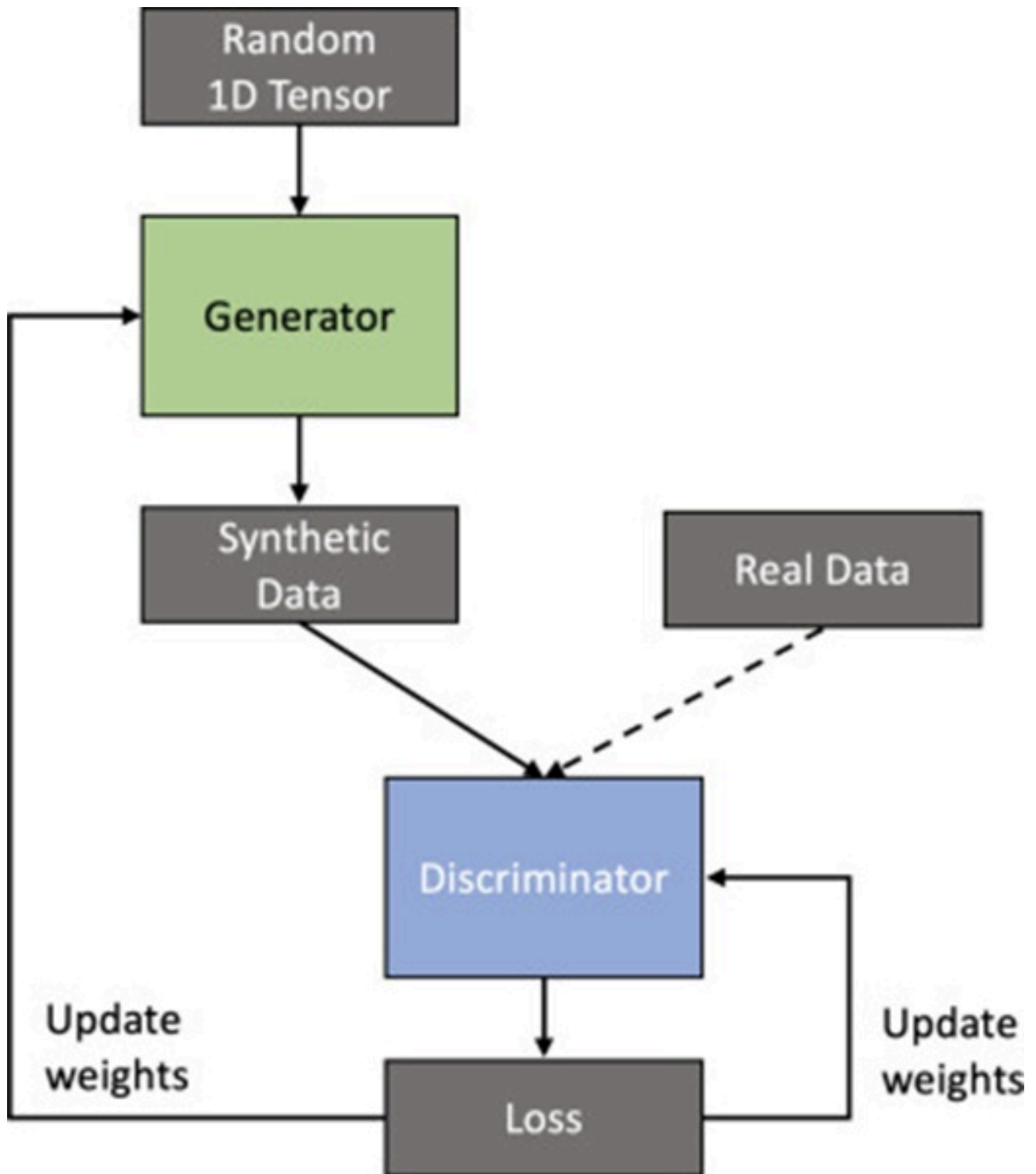


Figure 2. Principal sketch of a GAN.

In the binary classification setting, the cross-entropy loss for the i^{th} observation can be expressed in general form as:

$$S(y_i, \hat{y}_i) = -[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)], \quad (1)$$

where y denotes the true label ($y = 1$ for real data, $y = 0$ for fake) and $\hat{y}$ represents the predicted probability assigned by the discriminator.

The generator and discriminator pursue opposing objectives: the discriminator seeks to correctly label samples as real or generated. In contrast, the generator aims to produce outputs that cause the discriminator to misclassify them as real.

For a batch of observations, the average loss to be minimized by the discriminator and generator, respectively, can be formulated as follows:

Discriminator loss minimization

$$\min_D \left[- \left(\mathbb{E}_{x \sim p_{data}} [\log (D(x))] + \mathbb{E}_{z \sim p_z} [\log (1 - D(G(z)))] \right) \right] \quad (2)$$

Generator loss minimization

$$\min_G \left[- \left(\mathbb{E}_{z \sim p_z} [\log (D(G(z)))] \right) \right] \quad (3)$$

The global optimum is achieved when $p_g = p_{data}$, at which point the discriminator is no longer able to differentiate between the two distributions and outputs $(x) = 0.5$ for all inputs [46]. When substituted into the discriminator's loss function, this equilibrium yields an optimal loss value of $-\log(0.5) - \log(0.5) \approx 0.693$ upon convergence. In addition, the architecture exhibits the following characteristics:

- The generator learns to approximate the real data distribution without ever directly accessing the actual samples, doing so solely through the gradient information provided during backpropagation. The training dynamics resemble a zero-sum game: any improvement in the generator's performance (i.e., reduction in its loss) necessarily increases the discriminator's loss, and vice versa. As the generator improves at producing synthetic samples that fool the discriminator, the discriminator simultaneously refines its ability to detect fakes, prompting the generator to produce even more convincing outputs.

Training of the generator and discriminator occurs simultaneously rather than in alternation. Within each training iteration, both networks are updated. The generator's weights are adjusted while the discriminator's weights remain fixed during that step, because the gradients flow first through the discriminator before reaching the generator. Freezing the discriminator is essential because the two networks use different loss functions. Without this freeze, the generator could reduce its own loss simply by weakening the discriminator, rather than genuinely improving its approximation of the true data distribution p_{data} .

Conditional generative adversarial networks

Although Conditional Generative Adversarial Networks (CGANs) were introduced in the same year as the original GAN framework, they represent a significant advancement in GAN research. In the concluding remarks of their seminal paper, the authors of the first GAN [46] already suggested extending the architecture by incorporating an additional input shared between the generator and the discriminator.

This extra input functions as a conditioning signal that guides or constrains the generator's output. For instance, the condition might consist of a class label that supplies more specific information about the desired sample [12]. Such a label is typically encoded as a one-hot vector and fed into the network. The discriminator is then trained using both real and synthetic samples paired with their corresponding label vectors. This pairing is critical: without the matching condition, the discriminator would lack the information needed to evaluate whether a generated sample aligns with a particular label.

The same conditioning logic applies when processing real data during training. The loss functions for both the generator and the discriminator can be directly adapted from the unconditional GAN formulation, with the modification that the discriminator now estimates $D(x|y)$ —the probability of sample x being real given condition y —and the generator maps noise to data via $G(z|y)$, producing samples from p_g under condition y .

A distinctive element of this architecture is the concatenated layer. This layer combines the label tensor with either the noise tensor (for the generator) or the real/synthetic data tensor (for the discriminator). Mathematically, this operation can be described as follows: Consider a latent vector $\vec{z} = (z_1, z_2, z_3, \dots, z_{100})$ and a one-hot label vector $\vec{x} = (x_1, x_2, x_3, \dots, x_{10})$ representing 10 mutually exclusive classes. Concatenating these two 1D tensors produces a combined 1D tensor $\vec{c} = (z_1, z_2, z_3, \dots, z_{100}, x_1, x_2, x_3, \dots, x_{10})$. The same principle extends to higher-dimensional tensors. For concatenation to be feasible, the tensors must share identical dimensions in all axes except the one along which they are joined.

The concept of a conditional generator enables the application of GANs to time series synthesis under specific constraints, such as conditioning on the values of preceding time steps, as demonstrated by Koochali et al. [9].

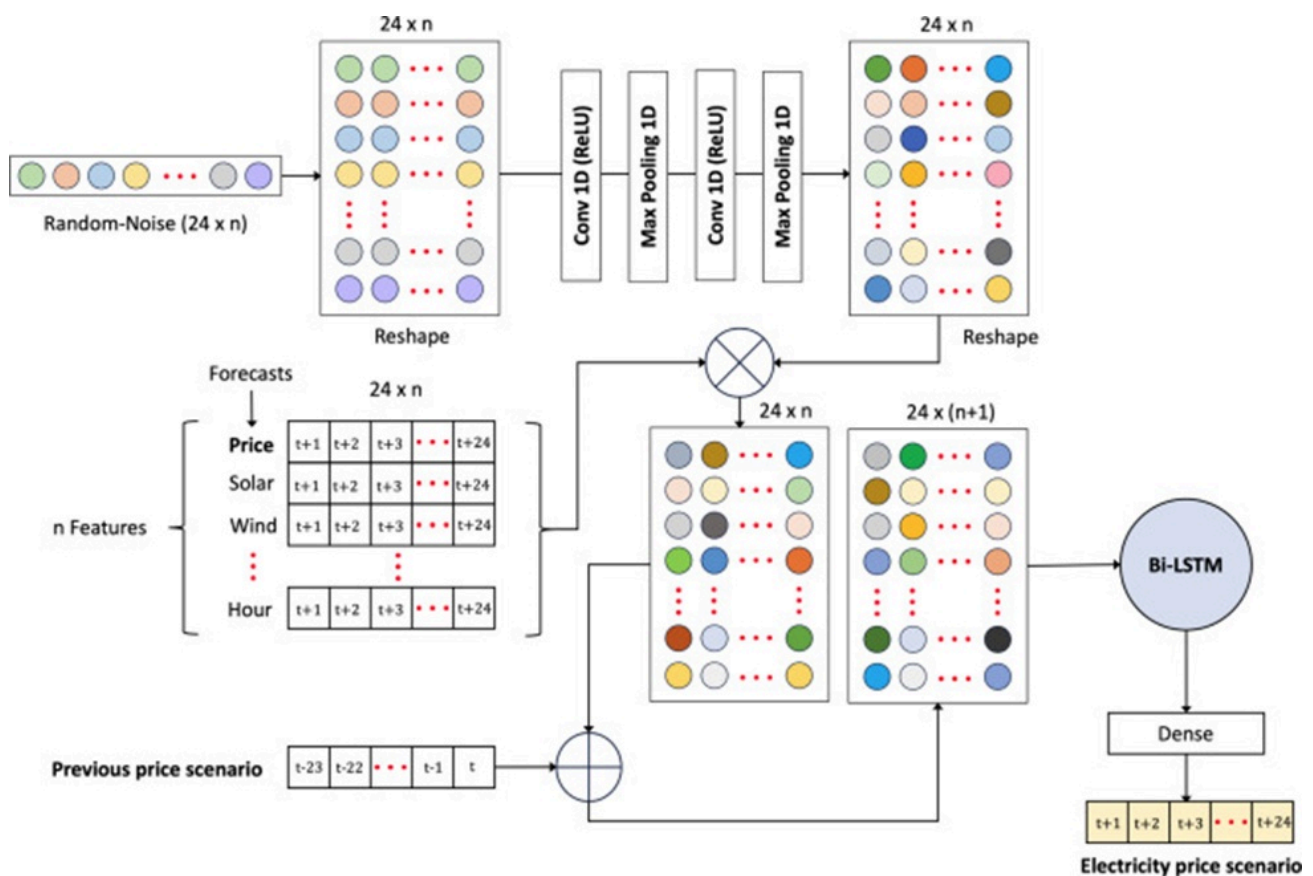


Figure 3. Model architecture of the generator of a TSS-CGAN.

Time series simulation conditional generative adversarial network

This subsection forms the central contribution of the present paper. We introduce a CGAN-based model architecture specifically designed for the probabilistic simulation of uncertain day-ahead electricity price scenarios. The foundational structure of this work draws inspiration from Huang et al. [11]. We extend and modify that architecture, along with its training procedure, to incorporate the following key adaptations:

- The generator produces a 24-dimensional output vector rather than a single scalar value. This design preserves the temporal ordering inherent in an electricity price profile and establishes interdependencies among the individual hourly time steps.
- In contrast to the approach in [11], where the target variable is duplicated within the feature matrix before the Bi-LSTM layer, we include the target variable only once. This adjustment ensures a more balanced influence from the remaining features.
- We implement a pre-training phase for the generator component (excluding the CNN layers and random noise input), treating it as a supervised regression task for 10,000 iterations to enhance initial prediction accuracy.
- We adjust the training dynamics by increasing the discriminator weight update frequency relative to the generator to a 2:1 ratio, thereby improving the discriminator's ability to accurately classify electricity price curves.

The precise architecture is detailed below. The generator produces a price scenario $\tilde{y}$ for the forthcoming 24 hours. Starting from a given time step t , the target prices are $\tilde{y}_{t+1}, \tilde{y}_{t+2}, \dots, \tilde{y}_{t+24}$. The conditioning input consists of the realized prices from the preceding 24 hours: $y_{t-23}, y_{t-22}, \dots, y_{t-1}, y_t$. The selection of relevant fundamental factors depends on the specific market under study. For the German market example used here, forecasts of wind and solar generation are particularly important. Additional factors—such as the availability of conventional generation capacity, energy commodity prices, and others—can also be included where appropriate. These fundamental variables pertain to the future period from $t + 1$ to $t + 24$ and are provided as forecasts. To further stabilize training, a point price forecast covering the same future interval is incorporated. Another essential input is the hour of the day for each data point, represented as an integer from 0 to 23. The price forecast, the fundamental forecasts, and the hour-of-day indicator are combined into a feature matrix, denoted as X .

$$X = \begin{bmatrix} Price_{t+1} & Price_{t+2} & \dots & Price_{t+24} \\ Solar_{t+1} & Solar_{t+2} & \dots & Solar_{t+24} \\ Wind_{t+1} & Wind_{t+2} & \dots & Wind_{t+24} \\ \vdots & \vdots & \ddots & \vdots \\ Hour_{t+1} & Hour_{t+2} & \dots & Hour_{t+24} \end{bmatrix}. \quad (4)$$

Thus, the generator can be expressed as a function $G(z, X, y)$ that takes a normally distributed random vector z , the historical price vector y (last 24 realized prices), and the feature matrix X to produce a simulated electricity price scenario $\tilde{y}$. The discriminator evaluates realism under identical conditioning inputs, so its output can be written as $(\hat{y}, X, y)$, where $\hat{y}$ denotes either a realized or generated price scenario. For a synthetic scenario, this becomes $(\tilde{y}, X, y)$ with $\tilde{y} = (z, X, y)$. Loss computation relies on binary cross-entropy. The architecture includes several distinctive elements, as shown in **Figure 3**.

The dimensionality of the latent space vector z remains an active research topic [47] and significantly affects the model's generalization performance. Given that the objective is to map high-dimensional data onto a low-dimensional distribution, the aim is to keep the noise dimension as compact as possible while retaining critical information. The ideal size of the latent vector varies widely across problem domains [47]. In the TSS-CGAN context, the dimension of z depends on the number of constrained features. After processing through two 1D convolutional layers and 1D max pooling layers, the vector z is reshaped to match the structure of the feature matrix X , enabling subsequent element-wise multiplication to integrate the latent representation with the conditioning information. This reshaping step is depicted in **Figure 3** for the generator and in **Figure 4** for the discriminator, occurring before the CNN processing stage. The element-wise multiplication—indicated by a circled $\times$ in both figures—establishes a direct arithmetic linkage between the properties

encoded in the latent space and the conditional features of the simulated scenario. The resulting matrix has shape $(24, n)$, where n denotes the number of features in X .

The realized prices from the previous 24 time steps are then concatenated (**Figures 3 and 4**) along the feature dimension to this matrix, yielding a final matrix of shape $(24, n + 1)$. This combined representation passes through a Bi-LSTM layer to capture temporal dependencies, followed by a Dense (fully connected) layer that produces the final 24-dimensional price vector. A defining characteristic of this architecture is the integration of a conditional feature matrix together with its inherent temporal organization.

The synergy between the Bi-LSTM layer and the 1D CNN layers is essential to the quality of the generated scenarios: the Bi-LSTM component extracts the temporal and sequential structure of the time series, while the 1D CNN layers capture spatial (i.e., feature-wise) patterns embedded in the latent space.

The discriminator follows a comparable structure, with the key distinction that the previously generated prices—rather than the latent space vector z —are fed into the 1D convolutional layer to extract spatial features before being multiplied element-wise with the feature matrix. An additional Bi-LSTM layer is employed to process the generated electricity price scenario alongside the sequence of historical realized prices. The outputs from both Bi-LSTM layers are then concatenated and passed through a Dense layer to produce the final result. Only the discriminator's output layer applies a logistic (sigmoid) activation function, while all preceding layers utilize the ReLU activation function. The ReLU (Rectified Linear Unit) activation function sets negative inputs to zero and leaves positive inputs unchanged, thereby introducing nonlinearity while maintaining computational efficiency. The complete architecture of the discriminator is illustrated in **Figure 4**.

The number of neurons in each layer is determined by the number of features under consideration and the number of time steps. For the number of filters in the two 1D convolutional layers and the configuration of the Bi-LSTM layers, we adopt the choices made by Huang *et al.* [11]. The first 1D convolutional layer contains 32 filters, the second contains 64 filters, the kernel size is set to 2 (including in the max pooling layer), and each Bi-LSTM layer comprises 24 units.

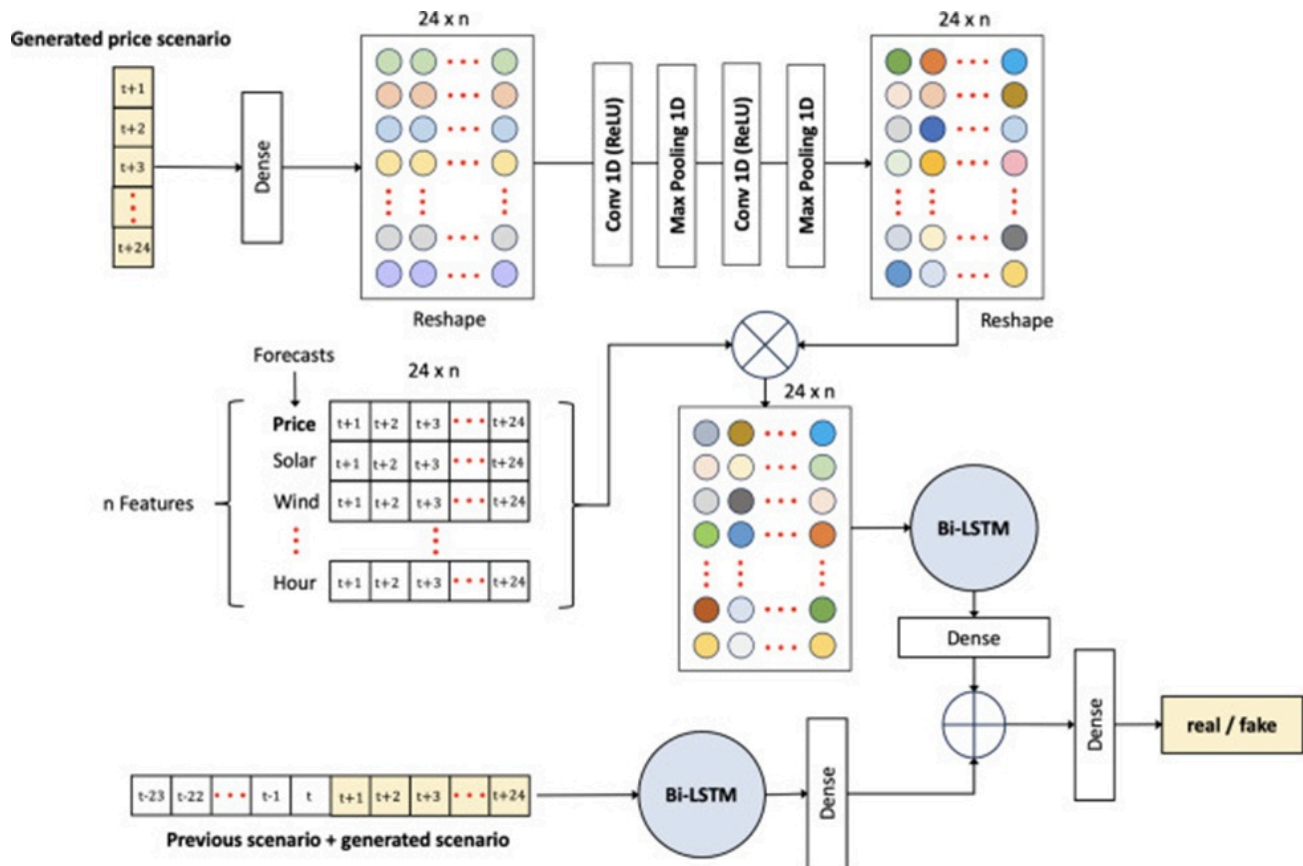


Figure 4. Model architecture of the discriminator of a TSS-CGAN.

Evaluation

The TSS-CGAN is employed here as a probabilistic model for generating electricity price scenarios. Evaluating such models can be approached from multiple perspectives. The realism of the simulated electricity price scenarios can be assessed qualitatively by examining distinctive patterns, such as intraday and weekly seasonality. In addition, a probabilistic model should produce a credible probability distribution over the simulated prices. This aspect is addressed through the quantitative evaluation of quantiles and the distribution of deviations between actual and generated prices. Finally, the model's responsiveness to electricity market-specific variables is tested through a sensitivity analysis that assesses whether it correctly captures and responds to the mechanisms governing electricity market behavior.

The implementation was performed on a workstation featuring an AMD Ryzen Threadripper PRO 7975WX 32-Core CPU, 128 GB of RAM, and two NVIDIA GeForce RTX 4090 GPUs. In practice, both model training and the sensitivity analysis are computationally intensive, each requiring approximately half a day on the described hardware configuration. All modeling was carried out using the TensorFlow/Keras library (version 2.13) in Python 3.10.

A dataset comprising 30,073 data points is available for training and evaluation. The observation period spans 3.5 years, from 17 June 2020 to 21 November 2023. The data include point forecasts of the hourly day-ahead electricity price (with a mean absolute error (MAE) of 10.12 €/MWh) as well as electricity market-specific fundamental variables (solar and wind generation, available capacity of gas, lignite, hard coal, and nuclear power plants, demand, and required residual load). All values pertain to the EPEX SPOT power exchange, specifically the Germany-Luxembourg market area. The corresponding realized day-ahead electricity prices are also included. An additional feature is the hour of the day, represented as an integer ranging from 0 to 23. In total, nine features are used:

- Price scenario forecast
- Previous day price scenario
- Solar energy forecast
- Wind energy forecast
- Residual load forecast
- Demand forecast
- Gas capacity forecast
- Coal capacity forecast
- Hour

Forecast data are not entirely consistent. For instance, occasional gaps appear and need careful handling during training. Whenever coal or gas capacity figures are absent, they are filled with the corresponding daily average—assuming the rest of the row is complete. This simplification introduces only a negligible error, since capacity numbers fluctuate much less than price predictions. Missing entries in the price forecasts are interpolated linearly. Complete days are discarded whenever any feature contains 12 or more missing hourly values, thereby preventing dataset distortion. In addition, the entire feature matrix is normalized using min–max normalization.

Training was performed on data from 17 June 2020 through 21 November 2022, with evaluation conducted on the hold-out period from 22 November 2022 to 21 November 2023. One defining trait of this test window is the unusually elevated market turbulence stemming from the pandemic aftermath and geopolitical tensions. The complete set of hyperparameters is listed below:

- Training iterations: 50,000
- Discriminator updates per generator iteration: 2
- Batch size: 500
- Optimizer: Adam (learning rate = 0.0001)
- Activation functions: ReLU throughout (sigmoid exclusively on discriminator output)
- Loss function: Binary cross-entropy
- Pre-training loss for generator: Mean squared error (MSE)

Learning rate, batch size, optimizer selection, and discriminator loss were directly taken from Huang *et al.* [11]. The total number of training iterations, however, was chosen empirically by monitoring when the discriminator loss leveled off, signaling convergence. Several discriminator update frequencies were tested, and a ratio of 2 discriminator steps per generator step yielded the best performance according to our chosen metrics.

Once training was completed, the generator produced 10,000 independent simulations over the test interval. Within each simulation run, 365 consecutive 24-hour price profiles are created. Every 24-hour block in the same simulation shares a

single fixed latent space vector throughout the entire test year. This consistent noise input ensures coherent mathematical relationships among all generated price paths over the full period.

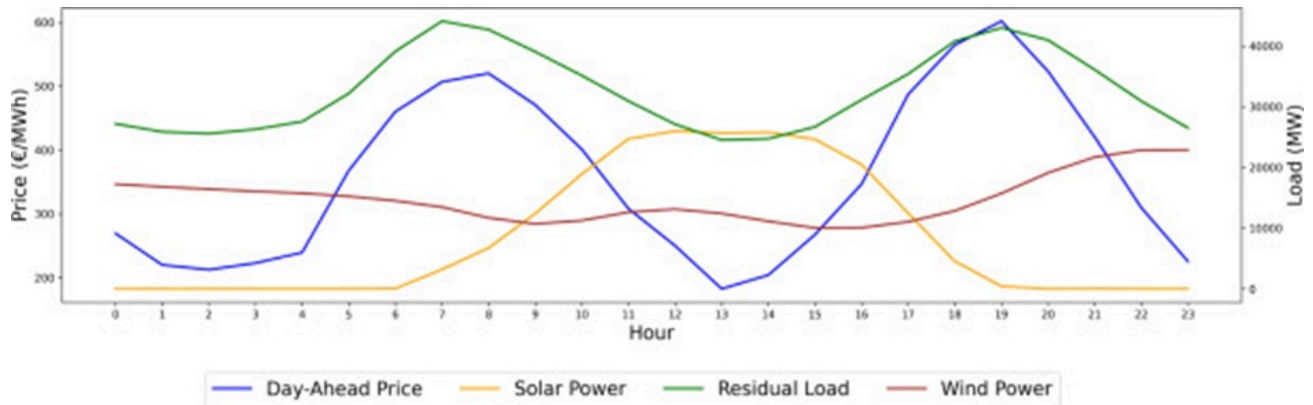


Figure 5. Characteristic day-ahead electricity price curve with a high share of renewable energy.

Qualitative analysis

Each day-ahead price scenario consists of exactly 24 hourly prices. The first value corresponds to the delivery window 00:00–00:59, while the last value covers 23:00–23:59. When these 24 points are connected, they form a daily price curve that reveals several well-known regularities. Elevated renewable infeed displaces conventional generation, thereby exerting downward pressure on prices through the merit-order mechanism. Days with exceptionally strong solar output typically display the familiar “camel hump” profile. Periods of low renewable contribution paired with elevated demand produce another recognizable shape. Moreover, weekends generally show lower short-term electricity consumption, resulting in noticeably lower overall prices.

Days dominated by renewable generation

When solar output is strong during the daytime, it substantially reduces the residual load that conventional sources must supply. On bright, clear days, this typically drives electricity prices sharply downward during the midday hours, even as prices climb higher in the early morning and late afternoon periods when residual demand remains elevated. A clear example appears in **Figure 5**, which displays the day-ahead price profile for Monday, 5 September 2022, and highlights exactly this dynamic. The x-axis marks the successive hours of the day. Prices (in €/MWh) are plotted against the left y-axis, while the right y-axis shows total renewable load (solar + wind) in MW together with residual load in MW.

Days with minimal renewable contribution

When renewables cover only a small fraction of total electricity needs, conventional thermal plants (mainly fossil-fuel-based) shoulder most of the supply burden. This situation produces price curves that respond more directly to demand fluctuations: prices rise steeply in the morning as load increases and fall more noticeably in the evening as load declines. In the absence of variable renewable injections, price movements remain comparatively smooth and stable during peak consumption hours. **Figure 6** captures this pattern on Friday, 9 December, 2022—a day when renewables played only a minor role in meeting demand—showing the flatter, more demand-driven behavior.

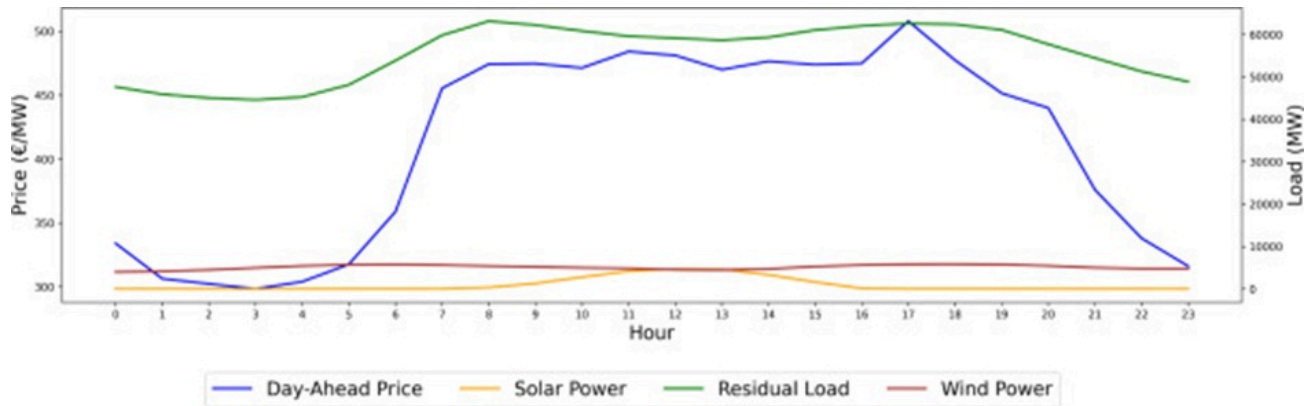


Figure 6. Characteristic day-ahead electricity price curve with low share of renewable energy.

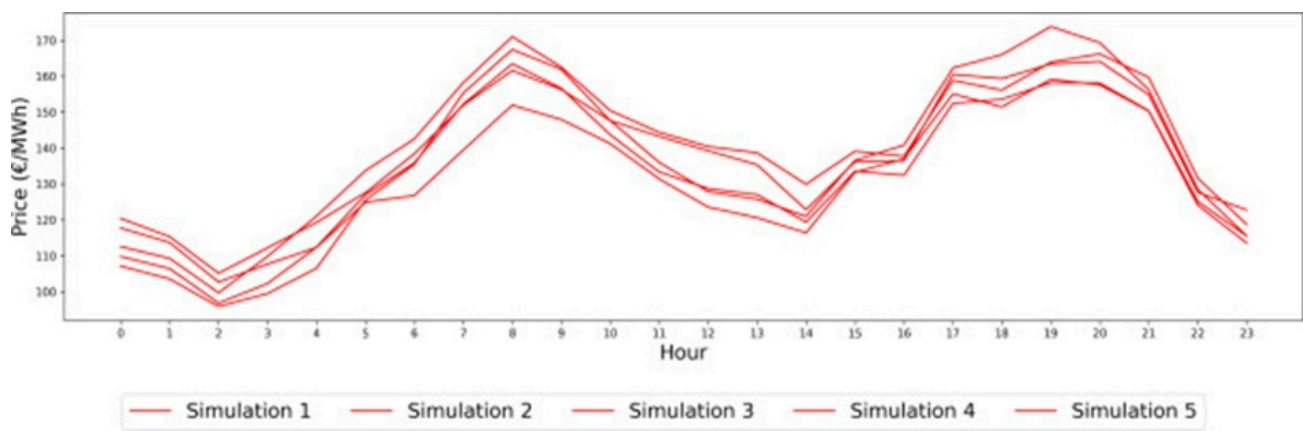


Figure 7. 5 Electricity price simulations.

Daily and weekly recurring patterns

Electricity prices are heavily shaped by the interplay between supply availability and consumption levels. Consequently, weekends almost always register significantly lower prices because industrial and commercial activity drops sharply. Overnight hours likewise show consistently lower demand and, therefore, lower prices than during daytime hours. **Figure 1** presents a complete weekly snapshot covering Monday, 28 November 2022, through Sunday, 4 December 2022.

The daily cycle stands out immediately: prices bottom out near midnight before climbing again through the morning and into the day. The figure's right-hand section emphasizes the weekend segment, with explicit markers at Friday 20:00 and Sunday 23:00. Both the price line and the residual load line decline visibly during these lower-demand weekend hours. Through qualitative inspection, we determine whether the model can reproduce these well-documented electricity-market regularities and whether the simulated curves retain the same fundamental supply–demand relationships observed in actual market data.

Quantitative analysis – forecast deviation

We quantify the probabilistic performance of the model using the Continuous Ranked Probability Score [48] (CRPS).

$$\text{CRPS}(F, x) = \int_{-\infty}^{\infty} (F(y) - 1\{y \geq x\})^2 dy \quad (5)$$

This score evaluates how accurately a predictive distribution F captures the true outcome. Provided F has a finite mean, the CRPS can be equivalently written in its negative form [49] (referred to here as CRPS*), defined by the expression:

$$\text{CRPS}^*(F, x) = \mathbb{E}_F |X - x| - \frac{1}{2} \mathbb{E}_F |X - X'| \quad (6)$$

Where X and X' are two independent samples drawn from F , and x stands for the observed realization (i.e., the actual prices recorded in the test set). One major practical advantage of CRPS* is that it returns a value in the same units as the target variable, making results directly interpretable. When applied to a deterministic point forecast, the second term disappears ($X = X'$), and CRPS* reduces exactly to the Mean Absolute Error (MAE). This equivalence allows meaningful head-to-head comparison between probabilistic and point-forecast models. The CRPS has become a standard metric both for probabilistic forecasting tasks in general [50] and for evaluating generative time-series models based on GANs in particular [17, 24].

Performance is benchmarked against three widely used probabilistic time-series models that can incorporate external covariates: DeepAR [51], Prophet [52], and the Temporal Fusion Transformer (TFT) [53].

Quantile ranges

By generating a sufficiently large number of simulations (10,000 or more), it becomes possible to compute quantile bands for each hour of the day. Figure 7 illustrates five example simulations overlaid for 22 February 2023. Importantly, because quantiles are calculated separately for each hour rather than across the full daily curve, simply connecting the 80% quantiles from every hour does not produce a single coherent simulated scenario. Instead, these quantile boundaries reveal the hourly probability structure—indicating the proportion of simulated prices that fall within particular ranges for that specific hour. In theory, exactly 60% of all simulated price values (not 60% of the complete daily curves) should lie between the 20% and 80% quantiles, since the quantiles are defined per hour.

This property is examined empirically using a hold-out test dataset. We generate 10,000 simulated price scenarios and then evaluate their quantile distributions using the procedure outlined in [1] to assess factor models on the German electricity market. Quantile intervals are constructed from the simulated outputs. Ideally, for example, precisely 20% of the observed (realized) test values should fall within the band bounded by the 0% and 20% quantiles. The same logic holds for the interval between the 20% and 40% quantiles, and so forth. Figure 8 presents a heatmap that visualizes the theoretical ideal outcome: in a perfectly calibrated model, the realized prices would be uniformly distributed across the quantile bands derived from the simulations.

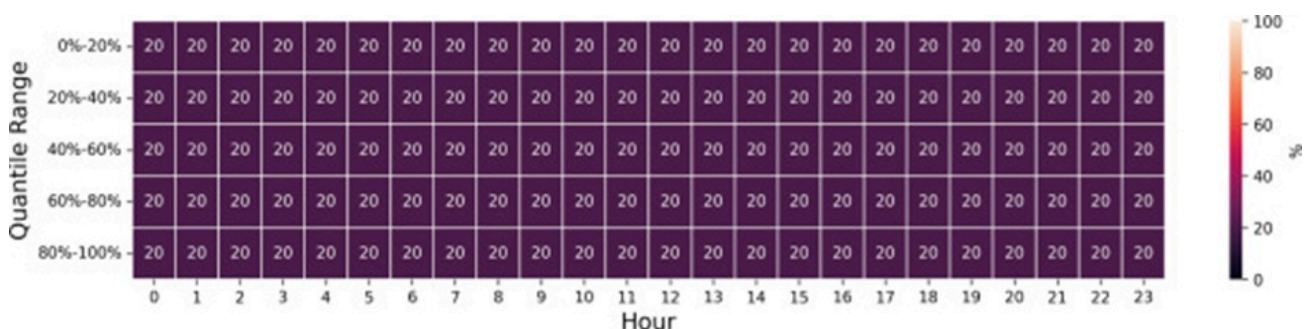


Figure 8. Heatmap of theoretical best case.

Distribution of difference

A further key quantitative measure in day-ahead electricity price modeling involves examining the distribution of deviations between simulated and actual prices during the test period, following the approach described in [14]. This difference distribution describes the likelihood of observing any given price deviation. In an unbiased and well-calibrated model, the mean of these differences should be zero, while the standard deviation should align with typical levels of electricity market volatility. Any noticeable skewness in the distribution would suggest a systematic tendency for the model to over- or under-predict prices.

Quantitative analysis – sensitivity

As demonstrated by numerous studies, conditioning GANs on rich feature matrices has proven highly effective. However, the direct impact of individual input features on the model's predictions cannot be inferred straightforwardly from the learned parameters. To address this, we conduct a sensitivity analysis by systematically perturbing single features and observing the resulting changes in output behavior. The goal is to verify whether the model appropriately responds to the core mechanisms that govern electricity market dynamics.

Key market drivers include fluctuations in renewable generation (which typically increase downside price risk while reducing upside risk), shifts in demand, changes in available conventional capacity, and variations in point price forecasts or preceding-day price patterns. In this analysis, we focus on perturbing the latter group of features. For each selected feature, we apply multiplicative adjustments of -30% , -20% , -10% , $+10\%$, $+20\%$, and $+30\%$, yielding six distinct scenarios per feature. For every perturbation level, 10,000 full simulations are generated. The resulting shift in the average simulated price serves as the primary indicator of the model's sensitivity to that feature. Figure 9 outlines the anticipated directional responses for each feature.

In real-world operation, an increase in renewable generation normally reduces residual load. To isolate individual feature effects in this controlled experiment, we deliberately refrain from simultaneously adjusting residual load when scaling renewable forecasts. This simplification is considered acceptable given the primary objective of cleanly measuring per-feature influence.

Feature	Increase Feature	Decrease Feature
Renewables FC	Higher Downside Risk	Higher Upside Risk
Fossil Fuels FC	Higher Downside Risk	Higher Upside Risk
Previous Day Price	Higher Upside Risk	Higher Downside Risk
Price FC	Higher Upside Risk	Higher Downside Risk
Demand FC	Higher Upside Risk	Higher Downside Risk

Figure 9. Expected behavior on variation of features.

Results and Discussion

The presentation of results follows the same sequence. At the close of this section, the findings are interpreted in relation to specific characteristics and mechanisms of electricity markets.

Qualitative analysis

The generated scenarios can be visualized as bands defined by selected quantiles. These can then be plotted as demonstrated in Figure 10, which displays simulated electricity price trajectories for Wednesday, 5 July, 2023. The overall price spectrum in the simulations spans roughly from -50 €/MWh to 300 €/MWh. Inspection of the red median line reveals

a clear pattern: prices climb during morning and afternoon hours but decline noticeably around midday and into the evening. This pattern aligns precisely with the dynamics previously shown and discussed in **Figure 5**. Both the median trajectory and the surrounding quantile bands successfully reproduce this characteristic behavior.

Further examination of the 98% prediction interval (which encompasses 98% of all simulated prices) reveals a substantial spread in the outputs. This wide dispersion effectively rules out mode collapse in the latent space vector in the presented example. Mode collapse is a training pathology in GANs in which the generator collapses into a single narrow mode, producing outputs regardless of the input noise. In the context of electricity price curves, this would manifest as nearly identical daily profiles across all simulations, irrespective of the latent vector or conditioning inputs. Mode collapse is a specific form of overfitting [54]. It can be mitigated, for instance, by carefully tuning the learning rates of the generator and discriminator, or by increasing the frequency of discriminator updates. Qualitatively, mode collapse in electricity price modeling is readily detectable by generating and overlaying many curves using diverse latent vectors: minimal variation or uniform appearance across different input signals is the issue.

Additionally, GAN convergence is indicated by the discriminator loss stabilizing near 0.6. Tracking the discriminator loss over training iterations, therefore, provides a useful diagnostic for assessing whether the model has reached equilibrium. The subsequent stage of qualitative evaluation involves comparing the simulated scenarios against both point forecasts and actual realized prices.

Days with a high share of renewable energy

Strong solar generation during daytime hours typically amplifies downward price pressure around midday. **Figure 10** illustrates simulated scenarios for a representative day exhibiting this effect. The blue curve representing realized prices remains comfortably within the 98% interval across all hours. The model clearly responds to elevated solar (yellow curve) and wind (brown curve) inputs by substantially lowering the lower boundary of the 98% interval. According to the simulations, negative prices fall well within the plausible range—an outcome observed in real market conditions under high renewable penetration.

Days with low share of renewable energy

When renewable sources provide only a minor portion of the total supply, conventional fossil-based generation assumes a larger role. **Figure 11** depicts the situation on Wednesday, 22 February 2023. The limited presence of variable renewables leads to noticeably flatter price profiles during peak-demand periods, as reflected in the narrow quantile bands. Both the median curve and the surrounding intervals display relatively stable behavior, closely mirroring the realized prices and the available forecasts. Over the course of the day, the wide extent of the 98% interval suggests a pronounced likelihood of elevated prices relative to depressed ones. Qualitatively, this upward price risk associated with low renewable shares appears consistent and plausible.

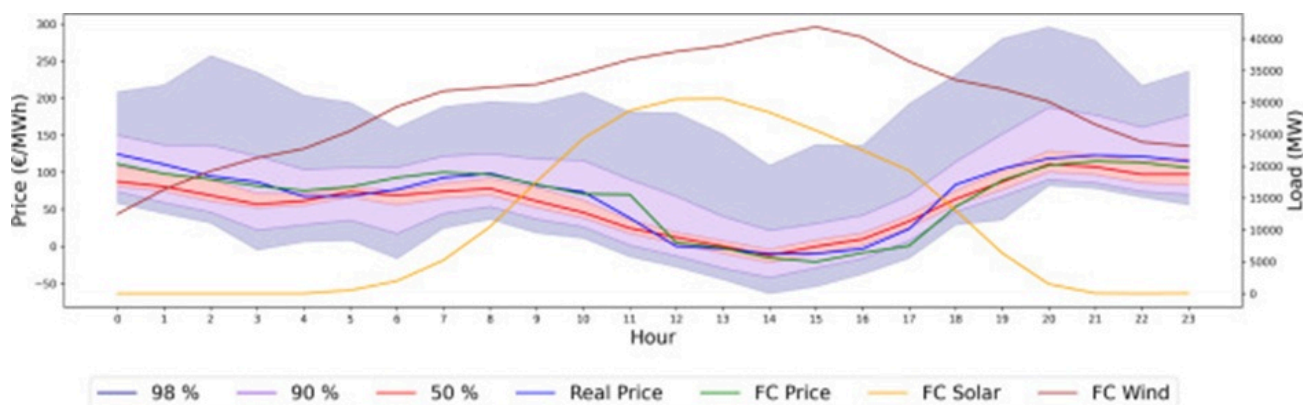


Figure 10. Simulations for a day with a high share of renewable energy.

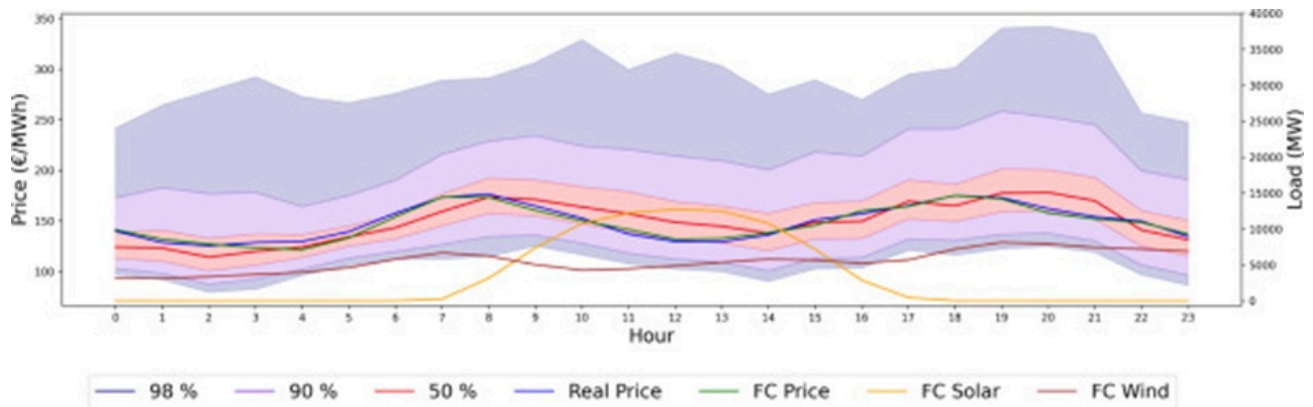


Figure 11. Simulations for a day with a low share of renewable energy.

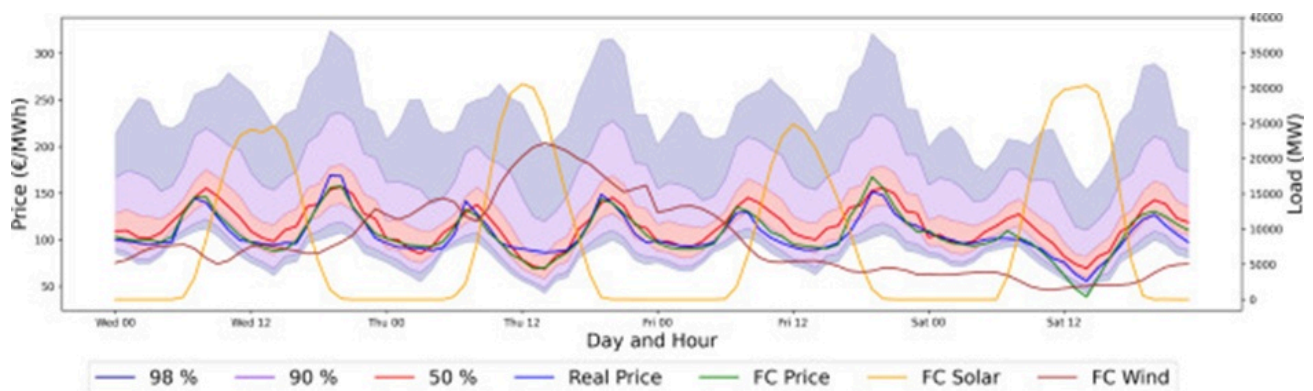


Figure 12. Simulations for multiple days.

Seasonalities in daily and weekly patterns

Extending the view across several consecutive days makes recurring day–night and weekday–weekend cycles clearly emerge. **Figure 12** presents simulated price scenarios covering Wednesday, 30 August 2023, through Saturday, 2 September 2023. The red median curve closely tracks the typical diurnal rhythm and shows pronounced sensitivity to solar and wind generation around midday. During noon hours, the lower boundary of the prediction intervals drops markedly, signaling elevated downside price risk. Both the median trajectory and the quantile bands indicate a distinctly higher probability of lower prices on Saturday than on weekdays, consistent with the well-known reduction in demand on weekends. Notably, every realized price throughout the period falls inside the 98% interval. The following subsection provides a quantitative assessment of how many actual observations land within the predicted quantile intervals.

Quantitative analysis – distribution of quantiles and deviations

Table 1 compiles the CRPS* scores obtained for the TSS-CGAN and the selected benchmark models. The proposed TSS-CGAN achieves the lowest (best) CRPS* value of 14.29 €/MWh. Prophet closely follows at 15.65 €/MWh, while TFT records 18.51 €/MWh. DeepAR shows the highest (worst) score at 28.49 €/MWh.

The TSS-CGAN harnesses the adversarial training framework of GANs to effectively learn intricate temporal relationships and faithfully represent the intrinsic uncertainty present in time series, proving especially valuable when data exhibit strong nonlinearities. By comparison, Prophet is purpose-built to model seasonal patterns and long-term trends, offering fast training and high interpretability—qualities that make it particularly well-suited to forecasting tasks dominated by clear

seasonality. TFT leverages attention mechanisms to selectively emphasize the most informative parts of the input history, enabling strong multi-horizon performance and explaining its clear advantage over DeepAR. DeepAR relies on autoregressive recurrent networks to generate predictions conditioned on historical observations and performs reliably on large collections of time series. Yet, it appears less capable of capturing the same level of complexity as TSS-CGAN.

Table 1. Results of CRPS in comparison to benchmark models.

Model	CRPS* in €/MWh
DeepAR	28.49
Prophet	15.65
TFT	18.51
TSS-CGAN	14.29

Although Prophet delivers quick, transparent forecasts, its relatively straightforward modeling assumptions can overlook subtle interactions that TSS-CGAN captures, rendering the latter more robust under volatile or structurally complex market conditions. TFT’s attention-based architecture enables it to surpass DeepAR on tasks that demand a refined understanding of sequential dependencies, consistent with earlier findings reporting up to 36% gains in multi-step load forecasting accuracy when using TFT [53]. DeepAR’s strength lies in its scalability across many parallel time series, which may favor it in certain high-volume applications. The markedly lower CRPS* obtained by TSS-CGAN demonstrates superior calibration of the predicted price distribution—an essential property for risk management and decision-making in electricity markets. The main drawback remains the substantially longer training time compared with Prophet’s near-instantaneous fitting. Consequently, while TSS-CGAN excels in distributional accuracy, Prophet may be the practical choice whenever rapid turnaround is prioritized.

Quantiles of the simulations

Quantile levels at 0%, 20%, 40%, 60%, 80%, and 100% are computed from the full set of simulated prices. In the ideal scenario, each of the five resulting intervals would contain exactly 20% of the realized test prices across all 24 hours. Figure 13 visualizes the outcome of this evaluation.

The empirical distribution of realized prices across these intervals reveals some deviations from uniformity. In particular, the model assigns greater probability to elevated price levels, as indicated by an over-representation of actual observations in the lowest quantile band (0%–20%). For certain hours—specifically contracts 5 and 10 through 13—up to 45% of realized values fall into this bottom interval, well above the expected 20%. Aggregating across all hours reinforces the same tendency toward higher simulated prices. The concentration of realized values in lower intervals during midday hours may partly stem from the pronounced price-depressing effect of solar generation, which appears stronger in reality than the model anticipates. Detailed results are presented in Table 2.

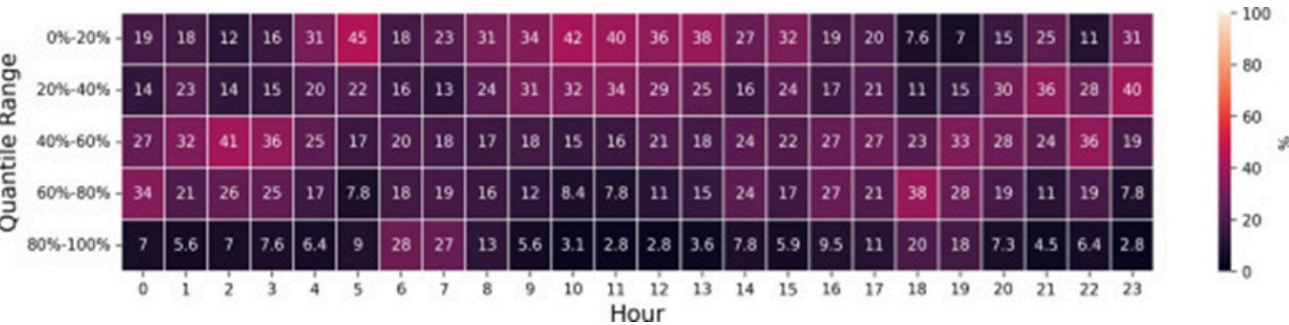


Figure 13. Frequency of real prices in specific intervals, hourly based.

Overall, the simulations exhibit a systematic bias toward higher price outcomes. One plausible explanation is that the training data included episodes of extreme volatility and very large positive price spikes. In contrast, the test period featured comparatively moderate price levels—as will be revisited in [Figure 14](#).

Table 2. Frequency of real prices in the test period in quantile ranges of 10,000 simulations.

Quantile range	% Real prices in quantile range
0%–20%	24
20%–40%	22
40%–60%	24
60%–80%	18
80%–100%	9

Distribution of difference

The likelihood of positive versus negative deviations can be effectively visualized through a probability distribution. [Figure 14](#) displays the distribution of differences between simulated and actual prices across the complete test period.

The shape closely approximates a Gaussian (normal) distribution, with probabilities tapering off symmetrically in both the positive and negative directions as the magnitude of deviation grows. A red vertical line marks the mean, which stands at +55 €/MWh. The median, shown as an orange vertical line, is located at +35 €/MWh. Dashed vertical lines indicate the standard deviations. Notably, the spread around the mode is asymmetric, with the positive tail extending farther than the negative one.

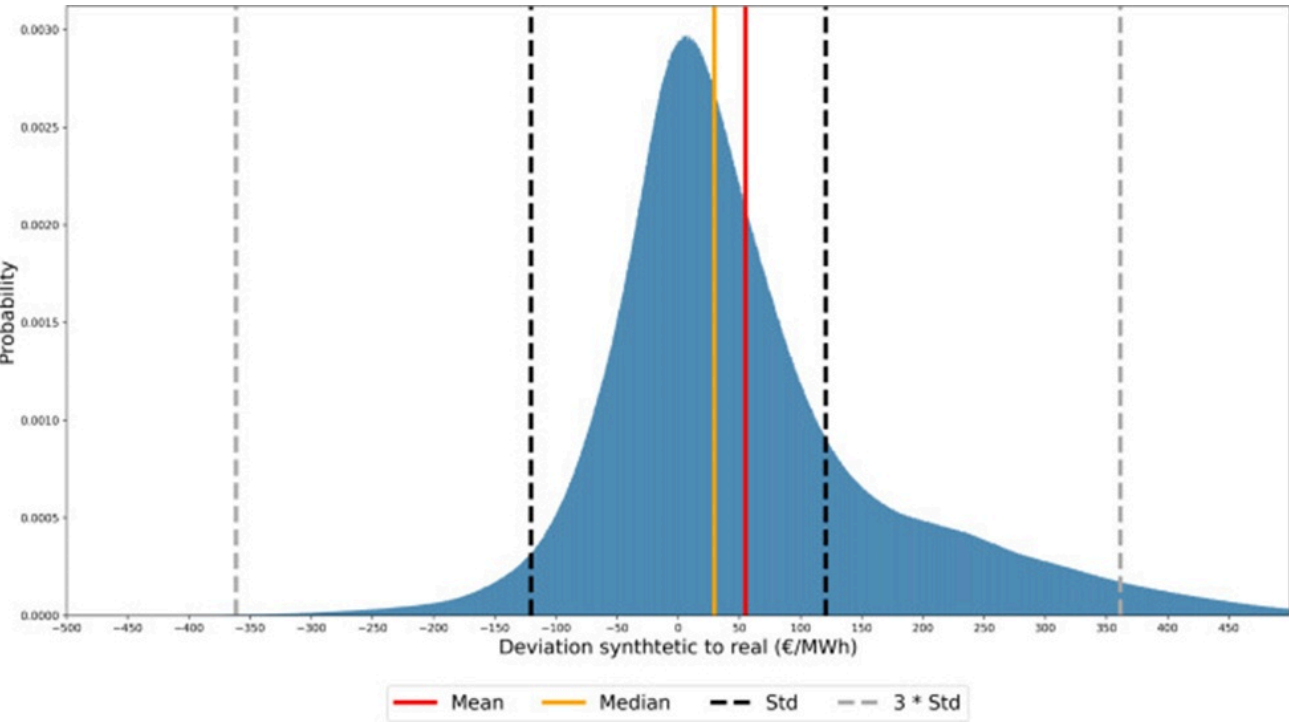


Figure 14. Deviation distribution of simulations from real prices.

Quantitative analysis – sensitivity

The final component of the quantitative assessment consists of a sensitivity analysis that systematically perturbs selected input features. Each chosen variable is scaled by 10%, 20%, and 30%. The outcomes of this analysis are summarized in **Table 3**. All reported figures represent the percentage change in average simulated price relative to the baseline (unperturbed) simulation.

In every tested scenario, the model shifts the average electricity price in the direction expected by logic (**Table 3**). It demonstrates clear sensitivity to changes in fundamental inputs. For instance, raising the renewable generation forecast by 30% results in an average price reduction of approximately 5%. The model shows even stronger responsiveness to adjustments in price-related features: increasing the point price forecast by 30% produces an average 14% rise in simulated prices. A comparable but somewhat milder effect occurs when scaling previous-day prices—here, a 30% increase leads to roughly 9% higher average scenarios. This greater influence of direct price forecasts is reasonable, since those forecasts pertain exactly to the simulated future period.

In contrast, the model exhibits near-neutral behavior with respect to demand variations: a 30% increase in forecasted demand results in only about a 1% increase in average prices over the evaluated period. Perturbations to fossil-fuel capacity values elicit a consistent but weak positive price response—an increase in available capacity leads to slightly elevated simulated prices. Whether this counterintuitive direction accurately reflects real-market dynamics is debatable. One could argue that greater capacity availability might push prices down by alleviating scarcity; alternatively, higher capacities could enable more expensive marginal units to enter the merit order, thereby exerting upward pressure. Regardless, the model's limited sensitivity to changes in fossil capacity and demand can be viewed as advantageous, as it suggests robustness against potential systematic biases or data inaccuracies in those variables.

Table 3. Results of sensitivity analysis.

Feature	Change	10%	20%	30%
Demand FC	+	0.07	0.56	0.66
	-	-0.51	-0.71	-1.31
Fossil Fuel FC	+	0.56	0.92	1.41
	-	-0.63	-1.39	-2.07
Prev Day Price	+	3.12	6.14	9.43
	-	-3.65	-6.93	-9.35
Price FC	+	4.70	8.88	13.85
	-	-4.93	-10.09	-14.53
Renewables FC	+	-1.21	-3.09	-4.65
	-	1.14	2.68	4.19

Further analysis

The preceding findings consistently indicate a systematic overemphasis on higher price levels in the simulated distributions. Closer inspection of the target variable itself sheds light on the origin of this tendency. **Figure 15** compares the distributions of realized day-ahead prices during the training and test periods.

Both distributions are markedly right-skewed, yet the training data exhibit far more extreme positive outliers. The European energy market experienced severe disruption in 2022 due to geopolitical tensions, resulting in exceptionally high volatility and uncertainty. This instability manifested in extraordinarily elevated prices; for example, the wholesale day-ahead price on 29 August 2022, for the 19:00–20:00 delivery hour, reached 871 €/MWh—the highest level recorded since the inception of organized exchange trading [55]. Having been exposed primarily to this highly volatile training distribution, the model internalized a strong expectation of large upward price movements. Consequently, even during the comparatively calmer test period, the generated scenarios remain biased toward higher prices.

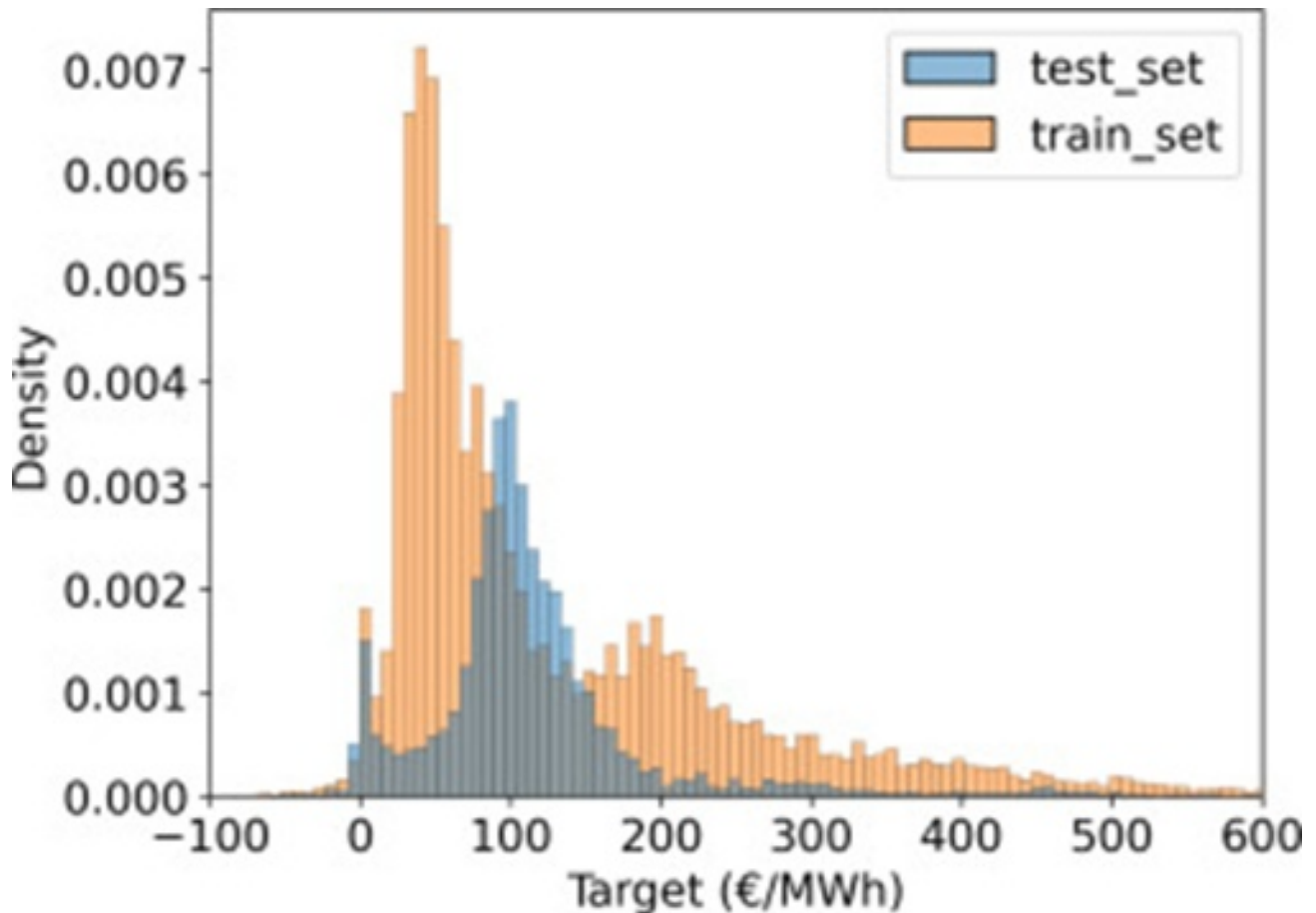


Figure 15. Distribution of target variable.

Conclusion

Strengths

The evaluation demonstrates that the generated electricity price scenarios fall plausibly within the spectrum of realistic market outcomes. Qualitatively, the model successfully captures the price-depressing impact of renewable energy sources and faithfully reproduces the characteristic intraday and weekly cyclical patterns. The sensitivity analysis further confirms the model's appropriate responsiveness to targeted changes in market drivers. Moreover, by leveraging the latent space vector, the TSS-CGAN produces a diverse array of plausible price paths, effectively preventing mode collapse.

Weaknesses

In the median trajectory, the model consistently underpredicts prices during early morning and late evening hours. The uppermost quantile band (80%–100%) should, in theory, encompass 20% of the realized test values, yet it contains only 9%. This same upward bias in the simulated distributions manifests in a skewed difference distribution and a predominance of negative deviations (simulated prices falling below realized prices). Additionally, the model's robustness to degraded input forecast quality remains untested, presenting a clear avenue for future investigation.

Outlook

This study demonstrates the construction of a Conditional Generative Adversarial Network tailored for generating day-ahead electricity price scenarios. The proposed TSS-CGAN is inherently adaptable: it can be trained on data from any electricity market and flexibly accommodates different sets of conditioning features, depending on the specific market context and the data available to the user.

Potential enhancements include replacing binary cross-entropy with a Wasserstein loss function incorporating gradient penalty (WGAN-GP) [56], following the approach taken in earlier relevant studies such as Demir *et al.* [20] and Lu *et al.* [21]. Such a modification could enhance training stability and improve generalization performance. Another promising direction is to increase architectural complexity—for instance, by adopting the TimeGAN framework introduced by Yoon *et al.* [13]. TimeGAN has already proven effective in several energy-domain applications, including day-ahead electricity price forecasting [22], synthetic load profile generation [57], and photovoltaic power output simulation [58].

Generative modeling of time series remains a relatively nascent research area, and its relevance—particularly for energy market applications—is likely to grow substantially in the years ahead.

Acknowledgements

This work was supported by the German Federal Ministry of Education and Research [grant 13FH587KX1] (FederatedForecasts).

Conflict of interest

None

Financial support

Viktor Walter reports that financial support was provided by the German Federal Ministry of Education and Research.

Ethics statement

None

Received: 28 Sep 2023

Revised: 05 Nov 2023

Accepted: 14 Dec 2023

Published online: 20 March 2024

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Hinderks WJ, Wagner A. Factor models in the German electricity market: Stylized facts, seasonality, and calibration. *Energy Econ.* 2020;85:104351.
- Bundesregierung. Wo steht Deutschland bei der Energiewende | Bundesregierung. 2023, <https://www.bundesregierung.de/breg-de/schwerpunkte/klimaschutz/faq-energiewende-2067498#:~:text=2021%20lag%20der%20Anteil%20der,weiter%20auf%20rund%2052%20Prozent>. Accessed 10 January 2024].
- Yu D, Ebadi AG, Jermisittiparsert K, Jabarullah NH, Vasiljeva MV, Nojavan S. Risk-constrained stochastic optimization of a concentrating solar power plant. *IEEE Trans Sustain Energy.* 2019;11(3):1464–72.
- Vithayasrichareon P, MacGill IF. A Monte Carlo based decision-support tool for assessing generation portfolios in future carbon constrained electricity industries. *Energy Policy.* 2012;41:374–92.
- Skantze PL, Ilic MD, Skantze PL, Ilic MD. A bid-based stochastic model for electricity prices. In: *Valuation, hedging and speculation in competitive electricity markets: A fundamental approach*. Springer; 2001. p. 61–111.
- Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative adversarial networks: An overview. *IEEE Signal Process Mag.* 2018;35(1):53–65.
- Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019. p. 4401–10.
- Chollet F. *Deep learning with Python*, NY, USA: Manning Publications; 2021;2.
- Koochali A, Schichtel P, Dengel A, Ahmed S. Probabilistic forecasting of sensory data with generative adversarial networks—forgan. *IEEE Access.* 2019;7:63868–80.
- Wiese M, Knobloch R, Korn R, Kretschmer P. Quant GANs: Deep generation of financial time series. *Quant Finance.* 2020;20(9):1419–40.
- Huang X, Li Q, Tai Y, Chen Z, Liu J, Shi J, et al. Time series forecasting for hourly photovoltaic power using conditional generative adversarial network and Bi-LSTM. *Energy.* 2022;246:123403.
- Mirza M, Osindero S. Conditional generative adversarial nets. 2014, arXiv preprint arXiv:1411.1784.
- Yoon J, Jarrett D, Van der Schaar M. Time-series generative adversarial networks. *Adv Neural Inf Process Syst.* 2019;32.
- Panapakidis IP, Dagoumas AS. Day-ahead electricity price forecasting via the application of artificial neural network based models. *Appl Energy.* 2016;172:132–51.
- Dudek G. Multilayer perceptron for GEFCom2014 probabilistic electricity price forecasting. *Int J Forecast.* 2016;32(3):1057–60.
- Tschora L, Pierre E, Plantevit M, Robardet C. Electricity price forecasting on the day-ahead market using machine learning. *Appl Energy.* 2022;313:118752.

Zúñiga G, Acuña G. Probabilistic multi-step time series forecasting using conditional generative adversarial networks. In: 2021 IEEE Latin American conference on computational intelligence. IEEE; 2021. p. 1–6.

Vuletić M, Prenzel F, Cucuringu M. Fin-gan: Forecasting and classifying financial time series via generative adversarial networks. *Quant Finance*. 2024;1–25.

Zhang Z, Wu M. Locational marginal price forecasting using convolutional longshort term memory-based generative adversarial network. In: 2021 IEEE Power and Energy Society General Meeting. IEEE; 2021. p. 1–5.

Demir S, Mincev K, Kok K, Paterakis NG. Data augmentation for time series regression: Applying transformations, autoencoders and adversarial networks to electricity price forecasting. *Appl Energy*. 2021;304:117695.

Lu X, Qiu J, Lei G, Zhu J. Scenarios modelling for forecasting day-ahead electricity prices: Case studies in Australia. *Appl Energy*. 2022;308:118296.

Lu X, Qiu J, Lei G, Zhu J. An interval prediction method for day-ahead electricity price in wholesale market considering weather factors. *IEEE Trans Power Syst*. 2023.

Bendaoud NMM, Farah N, Ahmed SB. Comparing generative adversarial networks architectures for electricity demand forecasting. *Energy Build*. 2021;247:111152.

Koochali A, Dengel A, Ahmed S. If you like it, gan it—probabilistic multivariate times series forecast with gan. *Eng Proc*. 2021;5(1):40.

Schwartz E, Smith JE. Short-term variations and long-term dynamics in commodity prices. *Manage Sci*. 2000;46(7):893–911.

Schnürch S, Wagner A. Electricity price forecasting with neural networks on epex order books. *Appl Math Finance*. 2020;27(3):189–206.

Singhal D, Swarup K. Electricity price forecasting using artificial neural networks. *Int J Electr Power Energy Syst*. 2011;33(3):550–5.

Yilmaz B, Laudagé C, Korn R, Desmettre S. Electricity GANs: Generative adversarial networks for electricity price scenario generation. *Commodities*. 2024;3(3):254–80.

Maria Hanif, Shahzad MK, Mehmood V, Saleem I. EPFG: Electricity price forecasting with enhanced GANS neural network. *IETE J Res*. 2023;69(9):6473–82.

Wang J, Chen Z. Factor-GAN: Enhancing stock price prediction and factor investment with generative adversarial networks. *Plos one*. 2024;19(6):e0306094.

Torres JF, Hadjout D, Sebaa A, Martínez-Álvarez F, Troncoso A. Deep learning for time series forecasting: A survey. *Big Data*. 2021;9(1):3–21.

Lara-Benítez P, Carranza-García M, Riquelme JC. An experimental review on deep learning architectures for time series forecasting. *Int J Neural Syst*. 2021;31(03):2130001.

Wan R, Mei S, Wang J, Liu M, Yang F. Multivariate temporal convolutional network: A deep neural networks approach for multivariate time series forecasting. *Electronics*. 2019;8(8):876.

Bu X, Wu Q, Zhou B, Li C. Hybrid short-term load forecasting using CGAN with CNN and semi-supervised regression. *Appl Energy*. 2023;338:120920.

Dong J, Chen J, Wu Q, Pan B, Liu G. Day-ahead prediction of wind power based on conditional generative adversarial network. In: 2021 IEEE Sustainable Power And Energy Conference. IEEE; 2021. p. 73–9.

Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* 1997;9(8):1735–80.

Zhang K, Zhong G, Dong J, Wang S, Wang Y. Stock market prediction based on generative adversarial network. *Procedia Comput Sci.* 2019;147:400–6.

Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM networks. In: 2005 IEEE international joint conference on neural networks, IEEE; 2005;4: 2047–52.

Siarni-Namini S, Tavakoli N, Namin AS. The performance of LSTM and BiLSTM in forecasting time series. In: 2019 IEEE international conference on big data. IEEE; 2019. p. 3285–92.

Garofolo JS, Lamel LF, Fisher WM, Fiscus JG, Pallett DS. DARPA TIMIT acousticphonetic continuous speech corpus CD-ROM. NIST speech disc 1-1.1. 93, NASA STI/Recon technical report n, 1993. p. 27403.

Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J. LSTM: A search space odyssey. *IEEE Trans Neural Netw Learn Syst.* 2016;28(10):2222–32.

Seabe PL, Moutsinga CRB, Pindza E. Forecasting cryptocurrency prices using LSTM, GRU, and bi-directional LSTM: A deep learning approach. *Fractal Fractional.* 2023;7(2):203.

da Silva DG, de Moura Meneses AA. Comparing long short-term memory (LSTM) and bidirectional LSTM deep neural networks for power consumption prediction. *Energy Rep.* 2023;10:3315–34.

Song J, Peng X, Yang Z, Qin G, Chen R, Wei P, et al. Wind power prediction under extreme weather conditions of low temperature based on TimeGAN and GWO-Bilstm. In: 2023 IEEE/IAS Industrial and Commercial Power Systems Asia. IEEE; 2023. p. 2319–25.

Géron A. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow, Sebastopol, CA, USA: O'Reilly Media, Inc; 2022;3.

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. *Adv Neural Inf Process Syst.* 2014;27.

Padala M, Das D, Gujar S. Effect of input noise dimension in GANs. In: Neural information processing: 28th international conference, ICONIP 2021, Sanur, Bali, Indonesia, proceedings, part III 28. Springer; 2021. p. 558–69.

Matheson JE, Winkler RL. Scoring rules for continuous probability distributions. *Manage Sci.* 1976;22(10):1087–96.

Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. *J Am Stat Assoc.* 2007;102(477):359–78.

Zamo M, Naveau P. Estimation of the continuous ranked probability score with limited information and applications to ensemble weather forecasts. *Math Geosci.* 2018;50(2):209–34.

Salinas D, Flunkert V, Gasthaus J, Januschowski T. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *Int J Forecast.* 2020;36(3):1181–91.

Taylor SJ, Letham B. Forecasting at scale. *Amer Statist.* 2018;72(1):37–45.

Lim B, Arik SÖ, Loeff N, Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int J Forecast.* 2021;37(4):1748–64.

Park S-W, Kim J-Y, Park J, Jung S-H, Sim C-B. How to train your pre-trained GAN models. *Appl Intell.* 2023;53(22):27001–26.

Göß S. Spot market prices for electricity break all records. 2022, . <https://blog.energybrainpool.com/en/spot-market-prices-for-electricity-breakall-records/>

Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville AC. Improved training of Wasserstein GANs. *Adv Neural Inf Process Syst*. 2017;30.

Asre S, Anwar A. Synthetic energy data generation using time variant generative adversarial network. *Electronics*. 2022;11(3):355.

Li Q, Zhang X, Ma T, Liu D, Wang H, Hu W. A multi-step ahead photovoltaic power forecasting model based on TimeGAN, soft DTW-based K-medoids clustering, and a CNN-GRU hybrid neural network. *Energy Rep*. 2022;8:10346–62.