Scenarios Modelling for Forecasting Day-Ahead Electricity Prices: Case Studies in Australia

Xin Lu[a], Jing Qiu[a], Gang Lei[b], Jianguo Zhu[a]

a. School of Electrical and Information Engineering, The University of Sydney, Darlington 2008, New South Wales, Australia

b. School of Electrical and Data Engineering, University of Technology Sydney, Ultimo 1994, New South Wales, Australia

Highlights:

1. A novel CTSGAN deep learning framework is first proposed and applied to model day-ahead electricity price scenarios.
2. The CTSGAN point forecasting model can eventually be transformed into a probabilistic forecasting model by enhancing the diversity of random input.
3. The CTSGAN probabilistic forecasting model can directly yield high-quality forecasting intervals with different coverage probabilities as a multi-objective forecasting model.
4. PSO-based optimal conditions selection method can increase the forecasting accuracy.

Abstract: Electricity prices in spot markets are volatile and can be affected by various factors, such as generation and demand, system contingencies, local weather patterns, bidding strategies of market participants, and uncertain renewable energy outputs. Because of these factors, electricity price forecasting is challenging. This paper proposes a scenario modeling approach to improve forecasting accuracy, conditioning time series generative adversarial networks on external factors. After data pre-processing and condition selection, a conditional TSGAN or CTSGAN is designed to forecast electricity prices. Wasserstein Distance, weights limitation, and RMSProp optimizer are used to ensure that the CTGAN training process is stable. By changing the dimensionality of random noise input, the point forecasting model can be transformed into a probabilistic forecasting model. For electricity price point forecasting, the proposed CTSGAN model has better accuracy and has better generalization ability than the TSGAN and other deep learning methods. For probabilistic forecasting, the proposed CTSGAN model can significantly improve the continuously ranked probability score and Winkler score. The effectiveness and superiority of the proposed CTSGAN forecasting model are verified by case studies.

Keywords: Generative adversarial networks; Point forecasting; Probabilistic forecasting; Electricity Price; Conditions

1. Introduction

Since the liberalization of electricity markets, the electricity price forecasting plays an essential role to guide the behaviors of participants for profit optimization, risk control, and stable operation [1-4], and the dynamics of electricity prices have become a complex phenomenon, which has brought significant challenges to price prediction [5]. In particular, day-ahead price forecasting has

attracted considerable attention [6-8]. For example, in virtual power plants, day-ahead price forecasting has been long advocated for improving the efficiency of arranging bidding strategies. However, electricity prices are highly volatile because of their strong dependence on multiple factors [9], such as electricity demand, business activities, temperature, and holidays. Moreover, the increasing integration of renewable energy sources, such as intermittent photovoltaic and wind power generation, has led to uncertain electricity generation and more significant fluctuations in the electricity prices. The electricity prices may also be affected by various potential factors, such as the large-scale application of artificial intelligence algorithms in electricity trading market [10] or even the personal bidding strategies of electricity traders. These factors are difficult to analyze quantitatively and have made electricity price forecasting significantly more challenging in recent years.

Nomenclature

GAN	Generative Adversarial Networks	x	Real scenarios
TSGAN	Time Series GAN	$\mathbf{P}_X$	Real historical scenario distribution
CTSGAN	Conditional TSGAN	$\theta^{(G)}$	Weights of generator
ML	Machine Learning	$\theta^{(D)}$	Weights of discriminator
ARIMA	Autoregressive Integrated Moving Average	L_G	Loss function of generator
ANN	Artificial Neural Network	L_D	Loss function of discriminator
ABC	Artificial Bee Colony	$e(\cdot)$	Embedding neural network
WNN	Wavelet Neural Network	$\theta^{(e)}$	Weights of embedding neural network
ELM	Extreme Learning Machine	$R(\cdot)$	Recovery neural network
SVM	Support Vector Machine	$\theta^{(R)}$	Weights of recovery neural network
PSO	Particle Swarm Optimization	x_t	Real scenarios at time t
DE	Differential Evolution	h_t	output of the embedding network
NSGA-II	Non-dominated Sorting Genetic Algorithm II	L_R	Loss function of reconstruction
DL	Deep Learning	L_U	Loss function of unsupervised
RNN	Recurrent Neural Network	L_S	Loss function of supervised
CNN	Convolutional Neural Network	$\|\cdot\|_2$	Euclidean norm
LSTM	Long Short Term Memory	λ , η	Balanced hyperparameters
GRU	Gated Recurrent Unit	$\mathbf{C}$	Price-related conditions
DNN	Deep Neural Network	Y_i	Actual prices at the time i
RMSE	Root Mean Square Error	y_i	Forecasting prices at the time i
MAE	Mean Absolute Error	n	Number of the prices in a day
LUBE	Lower Upper Bound Estimate	$F(\cdot)$	Cumulative distribution function
NEM	National Electricity Market	$I(\cdot)$	Indicator function
NSW	New South Wales	U_i	Upper bound of forecasting interval
VIC	Victoria	L_i	Low bound of forecasting interval
QLD	Queensland	δ_i	Width of forecasting interval
SA	South Australia	β , α	Confidence level
TAS	Tasmania	$\mathbf{C}_p$	Potential conditions
BOM	Bureau of Meteorology	$\mathbf{p}_{d-t}$	Daily historical electricity price of d-t day
ABS	Australian Bureau of Statistics	$\hat{\mathrm{p}}_d^{coal}$	Daily coal spot price (Newcastle)
AEMO	Australian Energy Market Operator	$\hat{\mathrm{p}}_d^{gas}$	Daily gas spot price (AEMO)
PCA	Principal Component Analysis	$\mathbf{L}_{d-t}$	Daily historical load profiles of d-t day
t-SNE	t-distributed Stochastic Neighborhood Embedding	$\hat{\mathbf{L}}_d$	Estimated load of the predicted day
MAPE	Mean Absolute Percentage Error	$\hat{\mathbf{T}}_d^{temp}$	Estimated temperatures of the predicted day
U2	U Statistics 2	$\hat{\mathrm{T}}_d^{hdd}$	Estimated heating degree day
CRPS	Continuously Ranked Probability Score	$\hat{\mathrm{T}}_d^{cdd}$	Estimated cooling degree day
WS	Winkler Score	$\hat{\mathbf{T}}_d^{cloud}$	Estimated duration of cloud
BPNN	Back Propagation Neural Network	$\hat{\mathbf{T}}_d^{sunshine}$	Estimated duration of sunshine
LASSO	Last Absolute Shrinkage and Selection Operator	$\hat{\mathbf{T}}_d^{shortwave}$	Estimated short wave radiation level
DM	Diebold Mariano	$\hat{\mathbf{T}}_d^{windspeed}$	Estimated wind speed
KDE	Kernel Density Estimate	$\hat{\mathbf{T}}_d^{winddirection}$	Estimated wind direction
QR	Quantile Regression	$\mathrm{D}_d^{week}, \mathrm{D}_d^{month}, \mathrm{D}_d^{holiday}$	Day types
QRRF	Quantile Regression Random Forest	$\mathrm{M}_d^{popu}, \mathrm{M}_d^{gdp}, \mathrm{M}_d^{cpi}$	Population, GDP and CPI of the NSW
$G(\cdot)$	Generator	$\mathrm{M}_d^{NECC}, \mathrm{M}_d^{NERC}, \mathrm{M}_d^{NEIC}$	Commercial, residential, and industrial electricity consumption
$D(\cdot)$	Discriminator	$\mathrm{M}_d^{AEG}, \mathrm{M}_d^{AREG}$	Total and renewable electricity generation in Australia

$\boldsymbol{Z}$	Random noise inputs	M_d^{NEG}, M_d^{NREG}	Total and renewable electricity generation in NSW
$\mathbf{P}_z$	Gaussian or uniform distribution		

1.1 Literature review

Extensive research has been done in the literature to improve the accuracy of price forecasting. Electricity price forecast models can be roughly classified into linear models and nonlinear models. Conventional linear models include autoregressive-based time series models, such as autoregressive moving average, autoregressive integrated moving average, fractionally integrated moving average, and autoregressive integrated moving average exogenous, and generalized autoregressive conditional heteroskedasticity. Dong et al. [11] proposed an ARIMA-based price forecasting model, combining empirical mode decomposition and seasonal adjustment, for electricity prices in Australia. Uniejewski et al. [12] introduced the seasonal component autoregressive model and quantile regression averaging model to forecast day-ahead spot price and gain further accuracy. However, the above methods are often criticized for their limited ability to capture the nonlinear behavior of electricity prices as electricity price changes are diverse. Recently, nonlinear methods represented by ML have been widely employed in price forecasting due to their outstanding performance in handling nonlinear problems. The ANN is the most utilized ML method, Shen et al. [13, 14] carried out an ABC combined WNN to achieve good forecasting performance. The ELM is another common type of ML method. Wan et al. [15] and Xiao et al. [16] modified the ELM with NSGA-II and DE-like algorithm to construct a forecasting model, respectively. The SVM is also often employed for prediction. Shrivastava et al. [17] carried out a novel forecasting model with PSO-tuned SVM and applied this model to Pennsylvania-New Jersey-Maryland interconnection day-ahead and real-time markets for validation. Because of the large number of factors involved in electricity prices, the conventional ML methods still cannot learn the characteristics of electricity prices well and have limited forecasting accuracy [18].

With the increasing computing capacity in recent years, DL has attracted worldwide research attention and has been applied to various fields, such as natural language processing [10, 19], process prediction [20], disease diagnosis [21], and autopilot [22, 23]. DL can analyze the deep and complex nonlinear relationship and then structure algorithms in layers to create an ANN that can learn and make intelligent decisions, such as RNN and CNN. RNN has been applied to forecast time series problems in many electricity-related fields [24, 25] and has shown good prediction performance due to the ability of the hidden neurons of the RNN to capture the short temporal dependency of the time series and then pass them to a previous layer. However, the previous research has also proven that RNN cannot capture the long temporal dependence during training because of the vanishing gradient problem [26]. For enabling RNNs to capture both long-term and short-term dependencies, LSTM was proposed by Hochreiter and Schmidhuber in 1997 [27]. There are three gates in LSTM, which can help control the information flow in or out of the memory block. GRU [28] is another form of RNN, in which the two gates, update gate and reset gate, work together to make the model converge rapidly. LSTM and GRU can address the vanishing gradient problem effectively for better forecasting performance. Simon et al. [29] developed several approaches of dynamic DL to predict the Spanish power market and found the seasonality periods and power demand as exogenous can help improve the accuracy. Peng et al. [30] used DE–LSTM to forecast electricity prices. The forecasting performance of the investigated method was verified in Australia, Germany/Austria, and France. Afrasiabi et al. [31] proposed a forecasting procedure consisting of CNN, GRU, and adaptive kernel density estimator to capture the probabilistic characterization of

real-time and day-ahead prices. Four different DL models, including DNN, LSTM-DNN, GRU-DNN, and CNN, were applied to forecast the spot electricity price, and it was found that DL could provide better forecasting than statistical models [32]. Though DL shows excellent fitting ability, it cannot still learn the whole temporal dynamics and may over-capture some useless features, leading to a poor generalization ability in electricity prices.

GAN is a new type of DL model that generally consists of a generator and a discriminator, both based on DL models. The discriminator in GAN identifies the difference between the generated and original scenarios to guarantee the predictive generalization ability and ensure that the generated scenarios are more realistic [33, 34]. In [35], a GAN model was introduced to model the financial time series in a data-driven manner. Chen et al. [36] developed a GAN-based wind forecasting model. The simulation results show that the unsupervised GAN model can generate a large number of scenarios, but the accuracy is not sufficiently high. While the wind power output is mainly influenced by physical factors such as wind speed and direction, electricity prices are related not only to the generation costs but also the supply and demand [37]. Wind and solar power output and load affect the electricity supply and demand and are a part of the electricity price forecast, making the spot electricity price forecast more complex. For example, in Jan. 2017, the volume-weighted weekly electricity price was A$508/MWh in QLD because of low solar generation. In Dec. 2019, low wind output drove the weekly electricity price to A$270/MWh in SA. Therefore, the GAN proposed in [35] cannot be used to predict electricity prices accurately. A good GAN model for time-series data should preserve the temporal dynamics, and the generated scenarios should respect the original distribution and relationship between variables across the whole period. Yoon [38, 39] introduced the latent space to GAN and proposed a TSGAN framework for synthesizing realistic scenarios, which integrates the versatility of unsupervised training with the control of supervised training and has a good predictive ability. Since electricity prices are related to multiple factors, relying on historical scenarios alone is insufficient. The TSGAN should be combined with a wide range of conditions to model the electricity prices.

In terms of forecasting objectives, electricity price forecasting can be divided into point forecasting and probabilistic forecasting. Point forecasting is generally deterministic forecasting that directly gives the exact outcome. For example, the deterministic outcome of day-ahead electricity price forecasting is to provide each half-hourly price of the next 24 hours, a total of 48 electricity price points. Most of the previous studies [11, 30, 32, 40] are about point forecasting of electricity prices and are concerned about the RMSE and MAE between the forecasted and actual data [41, 42]. However, some uncertainties are inevitable [43], usually due to incomplete data or some unanticipated events. In recognition of the inherent limitations of traditional point forecasting models, probabilistic forecasting of electricity prices is being used to compensate for the low accuracy of point forecasting and describe the uncertainties of electricity prices. Probabilistic forecasting can be presented in the form of quantile, interval, or density. A Pareto optimal price interval forecasting model was established in [15], combining NSGA-II and ELM to give confidence intervals directly by LUBE. Another form of probabilistic forecasting, quantile forecasting, was introduced in [8] to forecast distribution using quantile regression average. However, the above quantile and interval forecasting can only capture the probability distribution of the price individually in each period, and hard to integrate dependencies in the whole period [44]. A new scenario generation method for forecasting was proposed to model the variations effectively across periods in recent years. In [45], Karami et al. applied ARIMA to generate a massive number

of wind power output scenarios. The GAN model can directly give a massive number of scenarios and has been adopted to effectively forecast interval variations of wind power across different periods [46]. A new conditional GAN was proposed by Wang et al. [44] to model the probabilistic residuals of load forecasting. Due to the diversity of scenarios generated by the GAN model, it can be applied not only for point prediction but also for probabilistic prediction.

Many previous studies have focused on the electricity prices of the Australian NEM. The NEM operates one of the world's most extensive interconnected power systems and encompasses publicly and privately owned generators, transmission and distribution network providers, and traders [47, 48], and currently comprises five regions, namely NSW, VIC, QLD, SA, and TAS. As pointed out by Ignatieva [49], the regional electricity market may vary over time, and regional electricity prices interact through inter-regional transmission lines. The strongest dependence was exhibited between the NSW and QLD markets because they were well connected via inter-regional lines. In contrast, small dependence was found between markets without direct line connection, meaning that the electricity prices of QLD affect those of NSW more than those of TAS. Australia has four distinct seasons, with electricity prices showing different characteristics in different seasons. Uniejewski et al. [40] developed a seasonal component autoregressive forecasting model, consisting of a seasonal trend component and a stochastic component, for significant accuracy gains. Meanwhile, Australia is a world leader in the development of renewable energy generation, including wind and photovoltaic power. Large-scale grid integration of intermittent renewable energy generation has led to imbalances in electricity supply and demand, resulting in volatile electricity prices. By evaluating the merit order effect of five different levels of wind power generation penetration on electricity prices in NEM, Bell et al. [50] concluded that the increasing wind generation penetration decreased wholesale spot price in Australia. Forrest et al. [51] came to the same conclusion that wind power had a considerable impact on electricity spot prices and reduced the dispatch of emissions-intensive gas and brown coal generation. As a result of some policies, such as the introduction and abolition of carbon pricing, Australian electricity prices have gone through a process of rising and then falling. In [52], Nazifi et al. pointed out that carbon pricing would indeed be fully passed on to wholesale electricity spot prices resulting in higher electricity prices for consumers. Not only the external factors but also autogenous price fluctuations can cause changes in price trends. Higgs et al. [53] introduced three models to capture the high price volatility, strong mean-reversion, and frequent extreme price spikes in NEM, and concluded that the spot price exhibited stronger mean-reversion after a price spike than in the normal period, and the volatility was more than 14 times higher in spike periods than in normal periods. These factors lead to a large distribution of electricity prices in Australia and a high degree of difficulty in forecasting.

1.2 The knowledge gap and main contributions

The electricity prices forecasting is valuable for the participants to take part in the electricity market trading. The price of electricity is the result of a game between electricity suppliers and consumers, which presents a non-linear and difficult to forecast, and there are still some knowledge gaps as follows:

1) The conventional ML and DL models have successfully improved the forecasting accuracy, but hard to preserve temporal dynamics [54]. The forecasted electricity prices cannot respect the original distribution and relationship between variables across the whole period. On the other hand, GAN is difficult to use in industrial applications because the training process is difficult

to guarantee convergence.

2) Deterministic and indeterministic forecasting are usually separately performed using different methods, and there is a lack of a method that can be applied to both point and probabilistic forecasting with accurate performance.
3) Conventional point forecasting, constructed by the ML method, often leads to underfitting, while overfitting is always brought by the DL method. Whereas statistical-based probabilistic forecasting requires the assumption that the residuals satisfy some distribution, LUBE methods based on heuristic algorithms often fall into local optima.

In this paper, inspired by the conditional load model in [44] and TSGAN model, we propose a CTSGAN-based electricity price forecasting model and use NSW, Australia, as a forecasting example to verify the effectiveness of the proposed model under complex electricity price fluctuations. The major novel contributions of this work are as the following:

1) A novel CTSGAN deep learning framework is first proposed to capture the distribution of each period and preserve temporal variations, and a CTSGAN-based approach is designed to model electricity prices, effectively compensating for the gaps in conventional ML and DL models. In addition, Wasserstein Distance, weights limitation, and RMSProp optimizer are used to ensure the stability of the CTGAN training process.
2) An optimal condition selection method is proposed to extract more reasonable inputs of the CTSGAN model and make the forecasting results more accurate.
3) High realistic generated scenarios can be combined as the forecasting interval. When the noise dimensionality becomes large, the generated scenarios are diversified richly, and the point forecasting model is eventually transformed into a probabilistic forecasting model.
4) The CTSGAN-based deterministic electricity price point forecasting retains more original data features and shows better generalizations than other DL methods. Moreover, probabilistic forecasting can preserve the diversity and consistency of actual electricity prices and provides accurate intervals.

The rest of this paper is organized as follows: Section 2 introduces the methodology of the proposed CTSGAN model. Section 3 describes the forecasting steps and implementation details. Section 4 conducts the case studies in NSW for point forecasting and probabilistic forecasting, followed by the conclusion and discussions on the future research directions.

2. Methodologies

A good generative model should be able to capture not only the distribution of features at each time but also the complex dynamics of these variables across time. However, the existing generative methods that bring GAN into the sequential setting cannot adequately address the temporal correlations unique to time-series data. A new framework for generating realistic time-series that integrates the versatility of unsupervised training with the control of supervised training is proposed to simultaneously learn to encode features, generate representations, and iterate across time. The embedding network provides the latent space, and the adversarial network operates within this space. The latent dynamics of both real and synthetic data are synchronized through a supervised loss. Since electricity price forecasting involves various exogenous conditions, CTSGAN is proposed by using relevant conditions to modify TSGAN. In addition, the Wasserstein Distance is introduced to accelerate the training process and keep the training process stable. Moreover, two methods are applied to ensure the CTSGAN stability by (a) limiting the weights of the discriminator network

and (b) introducing RMSProp optimizer to replace Adam optimizer. This section introduces the proposed CTSGAN model and the steps of using it to forecast the day-ahead spot price scenarios.

2.1 The GAN model

The GAN model is mainly composed of two neural networks, i.e., the generator $G(\cdot)$ and the discriminator $D(\cdot)$. The generator's responsibility is to generate a massive amount of realistic scenarios with different random noise inputs, z, under a well-defined noise distribution, $z \sim \mathbf{P}_z$ (e.g., Gaussian distribution or uniform distribution) until the discriminator cannot distinguish them from the real scenarios, x. The responsibility of the discriminator, which is essentially a classifier, is to distinguish the generated scenarios and the real historical scenarios. Denote the generator neural network and discriminator neural network as $G(z;\theta^{(G)})$ and $D(x;\theta^{(D)})$, and the weights of neural networks in the generator and discriminator as $\theta^{(G)}$ and $\theta^{(D)}$, respectively. The ideal output of $G(z;\theta^{(G)})$ should follow the real historical data distribution, $X \sim \mathbf{P}_x$, after being well trained. The discriminator is trained to identify $\mathbf{P}_x$ from $\mathbf{P}_z$, and thus to maximize the difference between $D(x;\theta^{(D)})$ and $D(G(z;\theta^{(G)});\theta^{(D)})$. According to the objectives of the generator and the discriminator, loss functions L_G and L_D are formulated to train and optimize the weights of the neural networks of the generator and the discriminator, respectively. A small L_G indicates that the generator can synthesize realistic scenarios, which as if comes from the real historical scenarios. A small L_D reflects that the discriminator cannot identify whether the generated scenarios are from the historical scenarios. Define the L_G and L_D as:

$$L_G = -\mathbf{E}_Z\left[\log(1-D(G(z;\theta^{(G)});\theta^{(D)}))\right] \tag{1}$$

$$L_D = -\mathbf{E}_X\left[\log(D(x;\theta^{(D)}))\right] - \mathbf{E}_Z[\log(1-D(G(z;\theta^{(G)});\theta^{(D)}))\right] \tag{2}$$

where $\mathbf{E}_x$ and $\mathbf{E}_z$ denote the expected value over all real scenarios and the expected value over all random noise input to the generator, respectively. Thus, these two neural networks should be combined to construct the min-max optimization model:

$$\min_{\theta^{(G)}}\max_{\theta^{(D)}} \mathbf{E}_X\left[\log(D(x;\theta^{(D)}))\right] + \mathbf{E}_z[\log(1-D(G(z;\theta^{(G)});\theta^{(D)}))\right] \tag{3}$$

During the training process, the two neural networks in the generator and discriminator are optimized simultaneously, while both neural networks achieve the Nash equilibrium. When receiving random noise inputs, the generator can generate rather realistic scenarios that the discriminator cannot distinguish from the real scenarios. Moreover, by changing the random noise input, different scenarios can be generated.

2.2 The TSGAN model

A good GAN model for time-series data should preserve the temporal dynamics, and the generated scenarios can respect the original distribution and relationship between variables across the whole period. Yoon [38] proposed a TSGAN framework for synthesizing realistic scenarios, which integrates the versatility of unsupervised training with the control of supervised training. Unlike the basic GAN model, the TSGAN model consists of four neural network components: an embedding function, a recovery function, a generator, and a discriminator. A structure of the TSGAN model is shown in Fig.1.

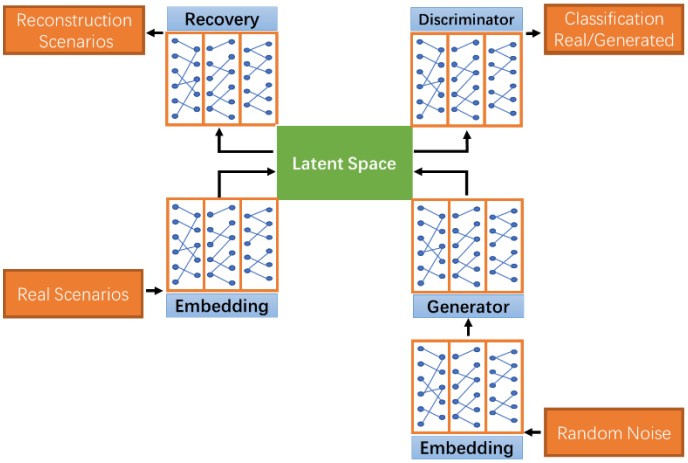


Fig. 1. Structure of the TSGAN model

Compared to the basic GAN, this model has two additional sections, the embedding section and the recovery section. These two sections provide the mapping between the feature and latent space, helping GAN learn the underlying temporal dynamics of the training scenarios via low dimensional representations. Denote the embedding neural network and the recovery neural network as $e(\cdot)$ and $R(\cdot)$, respectively. The real date input, x_t, and the output of the embedding network, h_t, at time t can be obtained respectively as the following:

$$h_t = e(h_{t-1}, x_t; \theta^{(e)}) \tag{4}$$

$$\tilde{x}_t = R(h_t; \theta^{(R)}) = R(e(h_{t-1}, x_t; \theta^{(e)}); \theta^{(R)}) \tag{5}$$

where $\theta^{(e)}$ and $\theta^{(R)}$ are the weights of two neural networks in the embedding and recovery networks, respectively. The embedding network and recovery network should enable accurate reconstruction of the original data from latent space, and therefore, the first objective is the reconstruction loss expressed as

$$L_R = \mathbf{E}_X \left[\sum_t \left(\left\| R(e(h_{t-1}, x_t; \theta^{(e)}); \theta^{(R)}) - x_t \right\|_2 \right) \right] \tag{6}$$

where $\|\cdot\|_2$ denotes the Euclidean norm. In TSGAN, the training period can be divided into two modes: the open-loop and closed-loop modes. In the open-loop mode, the generator and the discriminator work in the same way as the basic GAN. The second objective is the unsupervised loss, which is essentially the same as L_D in the basic GAN, as the following

$$\begin{aligned} L_U = {} & \mathbf{E}_X \left[\sum_t \left(\log(D(e(h_{t-1}, x_t; \theta^{(e)}); \theta^{(D)})) \right) \right] \\ & + \mathbf{E}_Z \left[\sum_t \left(\log(1 - D(G(e(h_{t-1}, z_t; \theta^{(e)}); \theta^{(G)}); \theta^{(D)})) \right) \right] \end{aligned} \tag{7}$$

To achieve the distribution of generated scenarios more efficiently, in the closed-loop mode, the generator receives the embedding network output to generate the next latent vector. The third objective is the supervised loss expressed as the following

$$L_S = \mathbf{E}_{X,Z} \left[\sum_t \left(\left\| e(h_{t-1}, x_t; \theta^{(e)}) - G(e(h_{t-1}, z_t; \theta^{(e)}); \theta^{(G)}) \right\|_2 \right) \right] \tag{8}$$

In the optimization period, the reconstruction loss and the supervised loss are trained first, and then the generator loss and discriminator loss will be trained. Combining the above three objectives, we can obtain the final optimization target as follows

$$\min_{\theta^{(e)},\theta^{(R)}} \left(\lambda L_S + L_R\right) \tag{9}$$

$$\min_{\theta^{(G)}} \left(\eta L_S + \max_{\theta^{(D)}} L_U\right) \tag{10}$$

where $\lambda \geq 0$ and $\eta \geq 0$ are hyperparameters that help balance both losses in each objective. Since TSGAN is not sensitive to these two hyperparameters [38], we set λ=1 and η=10. Compared with the objectives of basic GAN, L_S is a new optimization objective, which indicates that the embedding function not only serves to reduce the dimensions of the adversarial space but also facilitates the generator in learning underlying temporal dynamics.

2.3 The proposed CTSGAN model

The conventional electricity price forecasting using only historical electricity price data is generally challenging as the electricity prices are related to multiple factors. Therefore, modern electricity price forecasting methods consider a wide range of conditions. In this paper, the conditions related to electricity prices are added to the inputs of the generator and embedding networks. Combining the TSGAN and conditions, we can obtain a new model, known as CTSGAN. Fig.2 illustrates the structure of CTSGAN.

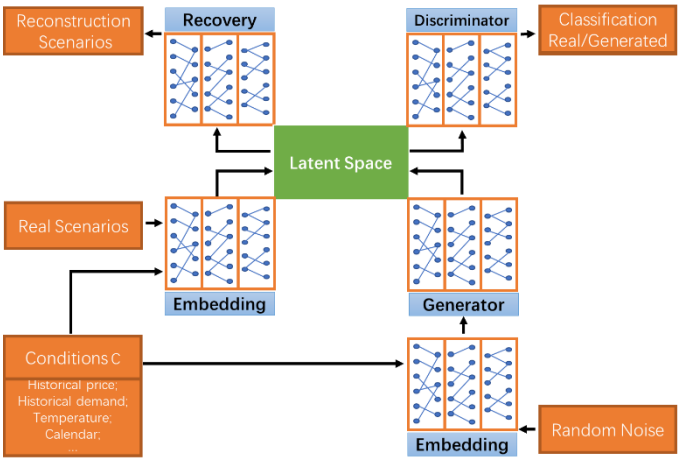


Fig. 2. Structure of the CTSGAN model

As shown, the inputs of generator networks contain not only random noise but also price-related conditions $\mathbf{C}$. On the other hand, the embedding network maps the real scenarios and conditions to the latent space. Denote the embedding neural network and recovery neural network as $e(\cdot)$ and $R(\cdot)$, respectively. The real date input, x_t, and the output of the embedding network, h_t, at time t can be obtained respectively as the following:

$$h_t = e(h_{t-1}, x_t|\mathbf{C}; \theta^{(e)}) \tag{11}$$

$$\tilde{x}_t = R(e(h_{t-1}, x_t \,|\, \mathbf{C}; \theta^{(e)}); \theta^{(R)}) \tag{12}$$

The loss functions of reconstruction, including the embedding and recovery neural networks, can be presented as the following:

$$L_R = \mathbf{E}_X \left[\sum_t \left(\left\| R(e(h_{t-1}, x_t|\mathbf{C}; \theta^{(e)}); \theta^{(R)}) - x_t|\mathbf{C} \right\|_2\right)\right] \tag{13}$$

The supervised and unsupervised losses in CTSGAN are:

$$L_S = \mathbf{E}_{X,Z}\left[\sum_t \left(\left\| e(h_{t-1}, x_t|\mathbf{C};\theta^{(e)}) - G(e(h_{t-1}, z_t|\mathbf{C};\theta^{(e)});\theta^{(G)})\right\|_2\right)\right] \tag{14}$$

$$\begin{aligned} L_U &= \mathbf{E}_X\left[\sum_t \left(\log(D(e(h_{t-1}, x_t|\mathbf{C};\theta^{(e)});\theta^{(D)}))\right)\right] \\ &+ \mathbf{E}_Z\left[\sum_t \left(\log(1 - D(G(e(h_{t-1}, z_t|\mathbf{C};\theta^{(e)});\theta^{(G)});\theta^{(D)}))\right)\right] \end{aligned} \tag{15}$$

Since the gradient of the logarithmic function in unsupervised loss (15) is very small in the early training stage, inspired by the Wasserstein GAN [55], the unsupervised loss function in the discriminator introduces the linear function instead of the logarithmic function to accelerate the training process and ensure training process stable. Thus, the unsupervised loss can be simplified as follows:

$$L_U = \mathbf{E}_X\left[\sum_t \left(D(e(h_{t-1}, x_t|\mathbf{C};\theta^{(e)});\theta^{(D)})\right)\right] - \mathbf{E}_Z\left[\sum_t \left(D(G(e(h_{t-1}, z_t|\mathbf{C};\theta^{(e)});\theta^{(G)});\theta^{(D)})\right)\right] \tag{16}$$

The unsupervised function can be interpreted as the dual of the Wasserstein distance (the Earth-Mover's distance). This distance between the generated and real scenarios measures the cost needed to transport $e(h_{t-1}, x_t|\mathbf{C};\theta^{(e)})$ to $G(e(h_{t-1}, z_t|\mathbf{C};\theta^{(e)})$. Another two methods are used to make the GAN stable: (a) after each iteration of discriminator training, all the weights of the neural network $\theta^{(D)}$ in the discriminator are limited to a specific range c as a clipping parameter, and (b) introducing the RMSProp optimizer to replace the Adam optimizer.

Algorithm 1 Scenarios Generation with CTSGAN

Input: Batch size m, number of iterations for each training period n, learning rate β, clipping parameter c
Output: $\theta^{(e)}$, $\theta^{(R)}$, $\theta^{(G)}$, $\theta^{(D)}$

#Update parameter for Embedding and Recovery
for *iteration* $= 1, 2, \ldots, n$ **do**
  Sample batch from historical data:
  $\{x^{(i)}\}_{i=1}^m \sim P_X$
  Update Embedding and Recovery nets using gradient descent:
  $g_{\theta^{(e)},\theta^{(R)}} \leftarrow \nabla_{\theta^{(e)},\theta^{(R)}}\left[\frac{1}{m}\sum_{t=1}^m \sqrt{(R(e(x^{(i)}|\mathbf{C})) - x^{(i)}|\mathbf{C})^2}\right]$
  $\theta^{(e)} \leftarrow \theta^{(e)} - \beta \cdot RMSProp(\theta^{(e)}, g_{\theta^{(e)}})$
  $\theta^{(R)} \leftarrow \theta^{(R)} - \beta \cdot RMSProp(\theta^{(R)}, g_{\theta^{(R)}})$
end for
#Update parameter for Generator only
for *iteration* $= 1, 2, \ldots, n$ **do**
  Sample batch from historical data:
  $\{x^{(i)}\}_{i=1}^m \sim P_X$
  Sample batch from random noise (Gaussian or Uniform):
  $\{z^{(i)}\}_{i=1}^m \sim P_Z$
  Update Generator net using gradient descent:
  $g_{\theta^{(G)}} \leftarrow \nabla_{\theta^{(G)}}\left[\frac{1}{m}\sum_{t=1}^m \sqrt{(e(x^{(i)}|\mathbf{C}) - G(e(z^{(i)}|\mathbf{C})))^2}\right]$
  $\theta^{(G)} \leftarrow \theta^{(G)} - \beta \cdot RMSProp(\theta^{(G)}, g_{\theta^{(G)}})$
end for
#Joint Training
for *iteration* $= 1, 2, \ldots, n$ **do**
  #Update parameter for Generator only (twice more than Discriminator)
  for *iterationGen* $= 1, 2$ **do**
    Sample batch from historical data:
    $\{x^{(i)}\}_{i=1}^m \sim P_X$
    Sample batch from random noise (Gaussian or Uniform):
    $\{z^{(i)}\}_{i=1}^m \sim P_Z$
    Update Generator net using gradient descent:
    $g_{\theta^{(G)}} \leftarrow \nabla_{\theta^{(G)}}\left[\frac{1}{m}\sum_{t=1}^m \left(\lambda\sqrt{(e(x^{(i)}|\mathbf{C}) - G(e(z^{(i)}|\mathbf{C})))^2} + (D(e(x^{(i)}|\mathbf{C})) - D(G(e(z^{(i)}|\mathbf{C}))))\right)\right]$
    $\theta^{(G)} \leftarrow \theta^{(G)} - \beta \cdot RMSProp(\theta^{(G)}, g_{\theta^{(G)}})$
    Update Embedding net using gradient descent:
    $g_{\theta^{(e)}} \leftarrow \nabla_{\theta^{(e)}}\left[\frac{1}{m}\sum_{t=1}^m \left(\eta\sqrt{(e(x^{(i)}|\mathbf{C}) - G(e(z^{(i)}|\mathbf{C})))^2} + \sqrt{(R(e(x^{(i)}|\mathbf{C})) - x^{(i)}|\mathbf{C})^2}\right)\right]$
    $\theta^{(e)} \leftarrow \theta^{(e)} - \beta \cdot RMSProp(\theta^{(e)}, g_{\theta^{(e)}})$
  end for
  #Update parameter for Discriminator only
  Sample batch from historical data:
  $\{x^{(i)}\}_{i=1}^m \sim P_X$
  Sample batch from random noise (Gaussian or Uniform):
  $\{z^{(i)}\}_{i=1}^m \sim P_Z$
  Update Discriminator nets using gradient descent:
  $g_{\theta^{(D)}} \leftarrow \nabla_{\theta^{(D)}}\left[\frac{1}{m}\sum_{t=1}^m (D(e(h_{t-1}, x_t|\mathbf{C})) - D(G(e(h_{t-1}, z_t|\mathbf{C}))))\right]$
  $\theta^{(D)} \leftarrow \theta^{(D)} - \beta \cdot RMSProp(\theta^{(D)}, g_{\theta^{(D)}})$
  $\theta^{(D)} \leftarrow clip(\theta^{(D)}, -c, c)$
end for

Compared with TSGAN, the proposed CTSGAN model can learn the underlying relationship between electricity prices and price-related conditions, making the generated scenarios relatively realistic. The pseudocode of the training algorithm of CTSGAN for electricity price forecasting is summarized in Algorithm 1.

3. The process for electricity price forecasting

Fig. 3 illustrates a flowchart to forecast electricity prices based on the proposed CTSGAN model. The forecasting steps and implementation details are discussed as follows.

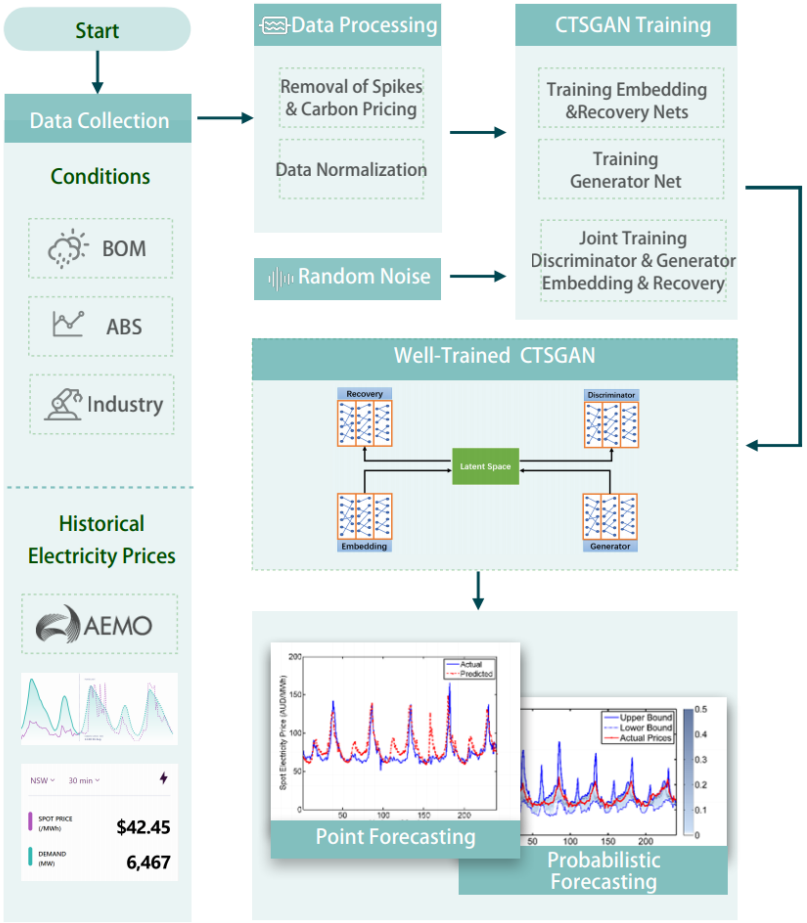


Fig. 3. Flowchart of the price forecasting based on CTSGAN model

3.1 The forecasting steps

The steps of forecasting the day-ahead electricity price based on CTSGAN are shown as follows:

Step 1. Data collection

The dataset of half-hourly electricity spot prices in NSW from 1 January 2000 is downloaded from the publicly available AEMO website https://aemo.com.au. The electricity price data for each month of the four calendar years, 2012, 2014, 2016, and 2018, and five days in each of the four seasons of winter (28 July to 1 August), spring (27 October to 31 October), summer (28 January to 1 February) and autumn (28 April to 2 May) in two financial years of 2018-2019 and 2019-2020 are selected as the test sets. The remaining data is set as the training data.

In addition to electricity prices, the remaining electricity market-related conditions, like half-

hourly electricity demand, are also obtained from the AEMO dataset. Electricity consumptions of different industrial components are from the annual report of AEMO. Besides, other conditions (such as weather, temperature, and economic index) are collected from the open-access websites of the BOM and ABS. The selection method of optimal conditions is described in 3.2.2.

Step 2. Data pre-processing

Removal of positive and negative spikes: The total historical electricity price dataset contains 332448 half-hourly electricity prices of 6926 days from 15 September 2001 to 31 August 2020. Since power failures, transmission line maintenance, and extreme weather can result in electricity price spikes and affect the accuracy of forecasting models; it is necessary to limit the electricity prices within a certain range [51]. In this paper, we set this range as [0,450] and the electricity price as A$450/MWh (symbol A$ represents the Australian Dollar) when it exceeds A$450/MWh and A$0/MWh when it is below A$0/MWh (negative electricity price is allowed in Australian Electricity Market). Counting through the dataset, we found 580 times of electricity prices over A$450/MWh and 61 times of electricity prices below A$0/MWh during this period, representing only 0.19% of the total data and thus having a limited impact on the forecasting model [56].

Removal of carbon pricing: Australia's carbon pricing scheme came into effect on 1 July 2012 and was repealed on 1 July 2014. Though this policy factor is not an input in the forecasting model, it impacts directly on the power generation from fossil fuels and renewable energy resources and thus significantly affects the electricity prices. To guarantee that the prediction model is not influenced by carbon pricing, the carbon tax should be removed from the dataset. The carbon pricing was A$23/t in the financial year of 2012-2013, and A$24/t in the financial year of 2013-2014. Considering that the carbon emissions factor in NSW is 0.9 t/MWh [57], one obtains that the carbon pricing included in the electricity prices in NSW is A$20.7/MWh in 2012-2013 and A$21.6/MWh in 2013-2014. This part should be removed during the training process and inserted back into the forecasts for the year of carbon price impact.

Data normalization: To fit the dataset to the forecasting model better, the data should be normalized to the range of $[0,1]$ with a min-max normalization method after removing the positive and negative spike prices and carbon pricing. The min-max normalization method is the linear transformation defined as the following:

$$p' = \frac{p - p_{\min}}{p_{\max} - p_{\min}} \tag{17}$$

where $p_{\min}$ and $p_{\max}$ are the maximum and minimum prices in the whole dataset of historical electricity prices, and p and p' the original and normalized prices, respectively.

Step 3. Optimizing CTSGAN weights

Before using CTSGAN to forecast the electricity day-ahead price, the weights of the embedding and recovery networks in CTSGAN should be optimized. The inputs of the CTSGAN are the pre-processing data of actual electricity prices, random noise, and conditions. The dimensionality of the random noise varies according to the forecasting objectives, with a smaller value chosen for point forecasting and a large value for probabilistic forecasting. The training processes for the point and probabilistic forecastings are independent of each other.

As shown in Algorithm 1, the training process can be divided into three stages. In the first stage, the optimization objectives are the weights of the embedding and recovery networks, which enable accurate reconstructions of the original electricity prices. In the second stage, the optimization objective is the weights of the generator. The generator receives the mapped data of actual electricity price computed by the embedding network and generates the next latent vector. The final stage is the joint training stage, in which the optimization objectives are the weights of the generator and discriminator. The generator is trained twice as many times as the discriminator is trained [39] to ensure the convergence of CTSGAN. When the maximum number of iterations is reached, the weights of the embedding and recovery networks, generator and discriminator have been optimized.

Step 4. Price forecasting and model evaluation

After the proposed CTSGAN forecasting model is well trained, the model can be used to roll over to forecast the day-ahead electricity price for the next day. For point forecasting, the dimensionality of the random noise is usually small, and therefore the variability of the forecasting is small. With the historical electricity prices, conditions, and 48-dimensional random noise as inputs to the forecasting model, the future 24-hour electricity prices can be obtained by just one forecast. However, for probabilistic forecasting, the dimensionality of the noise is usually large, and thus there is a rich diversity of forecasting results. By selecting different random noises that satisfy the Gaussian or Uniform distribution as random noise input of CTSGAN, multiple forecasted scenarios can be obtained and combined to form the probabilistic forecasting results. For both the point and probabilistic forecastings of electricity prices, different criteria are introduced to evaluate the performance of the proposed forecasting model as follows.

Visualization: The PCA is an unsupervised linear dimensionality reduction method by transforming the original set to a new set, known as the principal component. In this case, PCA tries to preserve the global structure of 48-dimensional day-ahead electricity price scenarios, and the local structures might get lost. The t-SNE is also an unsupervised nonlinear dimensionality reduction method. It embeds the points from a higher dimension to a lower dimension and tries to preserve the neighborhood of that point. Unlike the PCA, the t-SNE tries to preserve the local structure of data by minimizing the Kullback–Leibler divergence between the two distributions concerning the locations of the points in the map. In this case, it maps each 48-dimensional day-ahead electricity price scenario to a 2-dimensional point in such a way that similar scenarios are modeled by nearby points, and different scenarios are modeled by distant points with high probability.

Point forecasting criteria: Different error measures are employed as the criteria to evaluate the performance of day-ahead electricity price point forecasting models. The most used indicators are MAE, the MAPE, and RMSE, defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - y_i\right| \tag{18}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\left(Y_i - y_i\right)/Y_i\right| \tag{19}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - y_i\right)^2} \tag{20}$$

where Y_i and y_i denote the actual and forecasting electricity prices at the time i, respectively, and n is the number of the prices in a day. In the Australian electricity market, the settlement process operates on a 30-minute basis, and therefore, n is set to 48. Generally, a lower error measure indicates a better forecasting performance. As a measure of forecasting quality, the U2 proposed by Theil [58] is calculated by

$$U2 = \frac{\sqrt{\sum_{i=1}^{n}\left(\left(Y_{i+1} - y_{i+1}\right)/Y_i\right)^2}}{\sqrt{\sum_{i=1}^{n}\left(\left(Y_{i+1} - Y_i\right)/Y_i\right)^2}} \tag{21}$$

U2 is employed to compare the forecast quality between the proposed forecast results and naïve forecast results. The naïve forecast is a forecasting technique in which the actual electricity prices of the last moment are used as the forecast results of the current moment without adjusting them or attempting to establish causal factors. If U2=1, there is no difference between a naïve forecast and the proposed method; if U2<1, the proposed method is better than a naïve forecast; and if U2>1, the proposed method is worse than a naïve forecast.

Probabilistic forecasting criteria: The performance of the day-ahead electricity price probabilistic forecasting models should be evaluated from both uncertainty and variation should be considered. The uncertainty criteria contain reliability and sharpness. A considerable reliability value indicates that the forecasting scenarios should be able to cover the actual electricity prices, and a low sharpness value indicates that the forecasting value is close to the actual values at the same time. When both reliability and sharpness are met, the synthetic price scenarios are representative of the possible future realizations. The CRPS is employed to jointly evaluate the reliability and sharpness of scenarios forecasts by

$$CRPS(F, y_i) = \int_{-\infty}^{\infty} \left(F(y_i) - I(y_i - Y_i)\right)^2 dy_i \tag{22}$$

where $F(y_i)$ is the cumulative distribution function for forecasting prices scenarios value at the i-th time instant, $I(y_i - Y_i)$ represents the indicator to compare the forecasting and actual scenarios, $I(y_i - Y_i) = 1$ if $y_i > Y_i$, and otherwise, $I(y_i - Y_i) = 0$.

Not only the sharpness and reliability but also the confidence level needs to be taken into account. The WS is introduced to evaluate the interval forecasts as the following:

$$WS(y_i) = \begin{cases} \delta_i + 2 * \left(L_i - y_i\right)/\alpha, y_i \le L_i \\ \delta_i \qquad\qquad\quad , L_i < y_i < U_i \\ \delta_i + 2 * \left(y_i - U_i\right)/\alpha, y_i \ge U_i \end{cases} \tag{23}$$

where U_i and L_i denote the upper and low bounds of forecasting intervals of electricity prices, respectively, and $\delta_i = U_i - L_i$ is the width of the interval at i-th time instant. The upper and low bounds of intervals can be transformed from the scenarios if the confidence level $\beta \neq 1$, and $\alpha = 1 - \beta$. Then, $1 - \alpha/2$ and $\alpha/2$ are selected as the upper and low bounds of intervals. When the confidence level $\beta = 1$, the maximum and minimum values of each point of scenarios can be used as the upper and lower bounds for this point, respectively.

3.2 Implementation details

3.2.1 Selection and parameter setting of core NNs

The choice of core NNs in GAN is also a significant issue. Any NN can be used for GAN, even like the BPNN. However, when handling large datasets, it is best to employ DL models to guarantee that the training process does not collapse. For sequence forecasting, such as the electricity price forecasting in this paper, RNNs, including LSTM and GRU, are usually chosen. For model simplicity, LSTM is chosen as the core NNs of the embedding function, recovery function, generator, and discriminator in CTSGAN. The number of hidden layer units in LSTM is set to 100, and the number of training iterations is set to 10,000.

The parameters for the proposed CTSGAN are set as follows: the batch size is 7; the number for each training iteration is 10,000; the clipping parameter is 0.5; the learning rate is 0.02; the dimensionality of latent space is 100. The parameters for the TSGAN are set as the same as those of the proposed CTSGAN.

3.2.2 Selection of optimal conditions

Another task of preparation for forecasting day-ahead electricity prices is the selection of optimal conditions (features). Because electricity prices are related to many factors, by referring to the literature and relevant AEMO forecasting manual, data of various types of factors are downloaded from the public data websites of the Australian Government, including the Bureau of Statistics, the Bureau of Meteorology, and the Department of Industry, Science, Energy, and Resources, to obtain a dataset of potential conditions, $\mathbf{C}_p$, as the following:

$$\mathbf{C}_p = \begin{bmatrix} \mathbf{p}_{d-1}, \mathbf{p}_{d-2}, \mathbf{p}_{d-3}, \mathbf{p}_{d-7}, \mathbf{p}_{d-14}, \mathbf{p}_{d-21}, \hat{\mathrm{p}}_d^{coal}, \hat{\mathrm{p}}_d^{gas} \\ \hat{\mathbf{L}}_d, \mathbf{L}_{d-1}, \mathbf{L}_{d-2}, \mathbf{L}_{d-3}, \mathbf{L}_{d-7}, \mathbf{L}_{d-14}, \mathbf{L}_{d-21}, \\ \hat{\mathbf{T}}_d^{temp}, \hat{\mathbf{T}}_d^{cloud}, \hat{\mathbf{T}}_d^{sunshine}, \hat{\mathbf{T}}_d^{shortwave}, \hat{\mathbf{T}}_d^{windspeed}, \hat{\mathbf{T}}_d^{winddirection} \\ , \hat{\mathrm{T}}_d^{hdd}, \hat{\mathrm{T}}_d^{cdd}, \mathrm{D}_d^{week}, \mathrm{D}_d^{month}, \mathrm{D}_d^{holiday}, \mathrm{M}_d^{popu}, \mathrm{M}_d^{gdp}, \mathrm{M}_d^{cpi}, \\ \mathrm{M}_d^{AEG}, \mathrm{M}_d^{AREG}, \mathrm{M}_d^{NEG}, \mathrm{M}_d^{NREG}, \mathrm{M}_d^{NECC}, \mathrm{M}_d^{NERC}, \\ \mathrm{M}_d^{NEIC1}, \mathrm{M}_d^{NEIC2}, \mathrm{M}_d^{NEIC3}, \mathrm{M}_d^{NEIC4}, \mathrm{M}_d^{NEIC5}, \mathrm{M}_d^{NEIC6} \end{bmatrix} \tag{24}$$

where $\mathbf{p}_{d-t}$ denotes the daily historical electricity price of t days before the predicted day, and the length of daily electricity price is 48; $\hat{\mathrm{p}}_d^{coal}$ and $\hat{\mathrm{p}}_d^{gas}$ are the daily average coal spot price at the Newcastle port and natural gas price in AEMO, respectively; $\hat{\mathbf{L}}_d$ denotes the estimated load of the predicted day and $\mathbf{L}_{d-t}$ the daily historical load profiles of the t day before the predicted day; $\hat{\mathbf{T}}_d^{temp}$ denotes the estimated temperatures of the predicted day, and the length of $\hat{\mathbf{T}}_d^{temp}$ is 24, with which the heating degree day $\hat{\mathrm{T}}_d^{hdd}$ and cooling degree day $\hat{\mathrm{T}}_d^{cdd}$ can be calculated [59]; $\hat{\mathbf{T}}_d^{cloud}$, $\hat{\mathbf{T}}_d^{sunshine}$, $\hat{\mathbf{T}}_d^{shortwave}$, $\hat{\mathbf{T}}_d^{windspeed}$ and $\hat{\mathbf{T}}_d^{winddirection}$ are the estimated weather conditions, including the duration of cloud, duration of sunshine, short wave radiation level, wind speed, and wind direction, respectively; D_d^{week} and D_d^{month} are the day types (day of the week and month of the year), and $\mathrm{D}_d^{holiday}$ represents whether the predicted day is a holiday; M_d^{popu}, M_d^{gdp} and M_d^{cpi} denote the latest population, GDP and CPI of the NSW, respectively; M_d^{AEG}, M_d^{AREG}, M_d^{NEG} and M_d^{NREG} represent the total electricity generation and the renewable electricity generation in Australia and the state of NSW; M_d^{NECC}, M_d^{NERC} and M_d^{NEICi} denote the commercial, residential, and industrial components of the electricity consumption, respectively, and i in M_d^{NEICi} represents different industrial sectors; M_d^{NEIC1} represents the electricity consumption of agriculture, forestry, and fishing, M_d^{NEIC2} the electricity consumption of mining, M_d^{NEIC3} the electricity consumption of manufacturing, M_d^{NEIC4}

the electricity consumption of services of electricity, gas, water, and waste, M_d^{NEIC5} the electricity consumption of construction, and M_d^{NEIC6} the electricity consumption of transport, postal and warehousing. Although some data, such as the half-hourly electricity interregional transmission and electricity savings by policies, are unavailable, as their impacts on forecasting are limited, they are not discussed in depth in this paper.

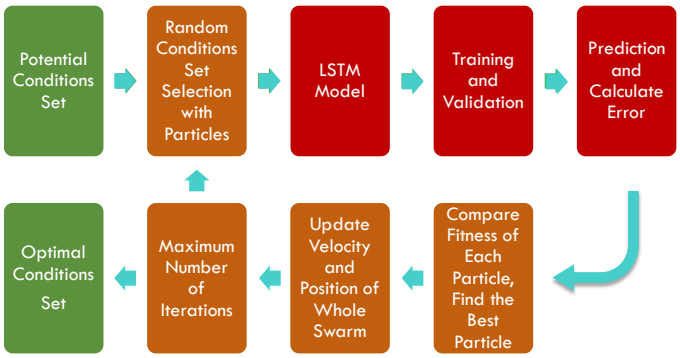


Fig. 4. Structure of the proposed model for selecting optimal conditions

With the potential set $\mathbf{C}_p$ and inspired by [60], we propose a PSO-LSTM condition selection model, as shown in Fig. 4. The PSO algorithm is applied to randomly choose a condition set from the potential set $\mathbf{C}_p$ for training the LSTM model. The well-trained LSTM model can then be used to forecast the day-ahead electricity prices. The PSO algorithm can be used to optimize the random condition set gradually by minimizing the error between the forecasting results and the actual prices. After 10,000 iterations, the optimal condition set of electricity prices, $\mathbf{C}$, can be obtained, as the following:

$$\mathbf{C}=\begin{bmatrix} \mathbf{p}_{d-1},\mathbf{p}_{d-2},\mathbf{p}_{d-7},\mathbf{p}_{d-14},\hat{\mathrm{p}}_d^{coal},\hat{\mathrm{p}}_d^{gas} \\ \hat{\mathbf{L}}_d,\mathbf{L}_{d-1},\hat{\mathbf{T}}_d^{temp},\hat{\mathrm{T}}_d^{hdd},\hat{\mathrm{T}}_d^{cdd},\mathrm{D}_d^{week}, \\ \mathrm{D}_d^{holiday},\mathrm{M}_d^{cpi},\mathrm{M}_d^{NREG},\mathrm{M}_d^{NERC},\mathrm{M}_d^{NEIC3} \end{bmatrix} \tag{25}$$

From the final optimal condition set, it is evident that the day-ahead price is related to several factors, including the historical price data, price of coal and gas, predicted and historical loads, temperate and heating or cooling degree day, CPI, renewable electricity generation in the NSW, and electricity consumption of residential and manufacturing in the NSW. The total dimension of the input of the CTSGAN is 322=48*6+24+10.

3.2.3 Dimensionality of random noise input

For day-ahead electricity price point forecasting, the desired outcome is low-diversity scenarios so that the dimensionality of random noise is set to 48, keeping it consistent with the number of inputs of the embedding network. For probabilistic prediction, it is necessary to guarantee the diversity of scenarios. Because of the large dimensionality of random noise [44], in the simulation, the process is repeated several times, and it is found that this value is best set to 100 for probabilistic forecasting. In addition, this property can be exploited to obtain different confidence levels and different prediction intervals by changing this value of dimensionality.

4. Case studies and analysis results

In this section, the proposed CTSGAN model is applied to point forecasting and probabilistic forecasting of electricity day-ahead prices in NSW, Australia.

4.1 Case 1: Electricity price point forecasting

As mentioned in Section 3.1, electricity price datasets from 15 September 2001 to 31 August 2020) are downloaded from AEMO. After data pre-processing, the price datasets are normalized to fit in [0,1] by using (17). The parameters are set as described in Section 3.2.1, and then the well-trained CTSGAN model can be employed to forecast electricity prices.

4.1.1 Experiment 1: Comparison with different forecasting models

To verify the point forecasting performance of the proposed CTSGAN model, we introduce four other neural networks, TSGAN, LSTM, GRU, and BPNN models, and two linear regression model, ARIMA and LASSO [61], for forecasting and comparison. Both the CTSGAN and TSGAN choose the LSTM as the core NN and the same parameter settings of LSTMs as described in Section 3.2.1. The GRU also has the same parameter as the LSTM. The number of layers and hidden units in BPNN [14] are set to 3 and 200, respectively, and the number of iterations is set to 50,000. The forecasts of electricity prices for 2012, 2014, 2016, and 2018 are conducted, respectively. Table 1 lists the results of four indicators, MAPE and U2 in p.u., and MAE and RMSE in A$/MWh, which can reflect the real world meaning well. Fig. 5 illustrates the average indicators from 2012 to 2018.

Table 1. Comparison of different point forecasting models for four years

Year	Indicators	CTSGAN	TSGAN	LSTM	GRU	BPNN	ARIMA	LASSO
2012	MAE	**4.82**	22.32	5.63	7.02	31.77	8.69	7.69
	RMSE	**6.17**	24.26	6.75	8.24	33.98	8.64	8.15
	MAPE	**0.1342**	0.7353	0.1656	0.1958	1.1626	0.2969	0.1961
	U2	**0.1617**	0.8665	0.2067	0.2196	1.3718	0.4333	0.2234
2014	MAE	**4.37**	23.85	7.88	5.54	35.91	9.32	8.10
	RMSE	**5.94**	25.83	9.27	6.93	37.94	9.59	7.11
	MAPE	**0.1765**	0.8827	0.3238	0.2156	1.3038	0.5236	0.3375
	U2	**0.1555**	0.9810	0.3108	0.1912	1.6995	0.3238	0.2186
2016	MAE	**16.20**	28.04	17.19	18.77	37.62	20.84	20.61
	RMSE	**20.52**	33.71	22.37	23.54	43.47	28.85	23.09
	MAPE	0.3069	0.4865	**0.2959**	0.3295	0.6681	0.4922	0.3081
	U2	0.1668	0.2512	**0.1555**	0.1626	0.3790	0.1937	0.1684
2018	MAE	**12.78**	39.92	17.82	19.98	54.77	24.12	20.84
	RMSE	**16.74**	43.83	22.59	25.92	58.01	37.89	23.31
	MAPE	**0.1570**	0.4793	0.2959	0.2527	0.6597	0.3044	0.3012
	U2	**0.1253**	0.3364	0.1960	0.2093	0.4405	0.3139	0.2330

In Table 1, the best or least value for each indicator in each year is marked in bold. As shown, overall, the proposed CTSGAN model gives the best performance in forecasting electricity prices, with a MAE of A$4.82/MWh and a RMSE of A$6.17/MWh, followed by the LSTM model, which also gives satisfactory forecasting performance in 2016 with a MAPE of 0.2959 and a U2 of 0.1555. The GRU model has slightly inferior forecasting ability to the LSTM model. The predictive ability of the LASSO model is similar to that of the GRU model for the prediction of 2016, with a RMSE of A$23.09/MWh (LASSO) and A$23.54/MWh (GRU), then followed by the ARIMA model. The TSGAN model has poor predictive results due to not considering the conditions, with all indicators larger than those of the GRU, LSTM, LASSO, and ARIMA models. This further demonstrates that it is not accurate to rely on the historical electricity price data alone to forecast electricity prices,

and it is necessary to consider the conditions, such as the weather, calendar, and demands. BPNN is the worst forecasting model, which is worse than the naïve forecasts for 2012 and 2014, with the U2 indicator greater than 1. From average indicators radar charts in Fig. 5, one can find that the proposed CTSGAN model is more effective than other models in terms of MAE, RMSE, MAPE, and U2. There is little difference in MAPE and U2 of LSTM, GRU, and LASSO, for which the forecasting accuracies fluctuate at the same level.

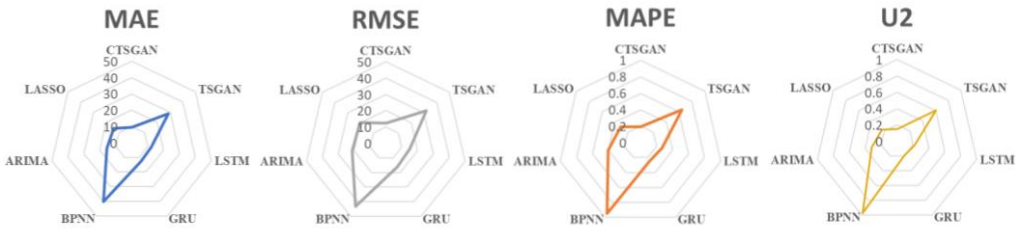
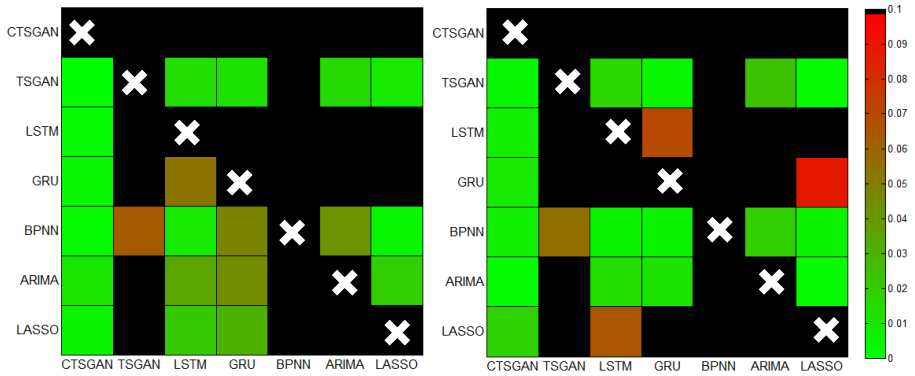


Fig. 5. Radar charts of average indicator based on seven models

However, the values in Table 1 can only be used to provide a model ranking, Uniejewski et al. [62] pointed out that DM [63] test can be introduced to verify whether the values in Table1 are statistically significant. Fig. 6 depicts the results of the DM test for MAE and RMSE in 2012. The color bar on the right represents the range of p-values: the closer they are to zero (→greener), the more significant the difference between the forecasts of a set on the X-axis (better) and those on the Y-axis (worse); otherwise, the closer they are to 0.1 (→redder or black), the less statistically significant the difference between the forecasts of models on the X-axis and Y-axis. For example, the second row in both Fig.6 is green except for one black square, which indicates that TSGAN gives poorer performance significantly than all other forecasting models except BPNN.



(a) (b)

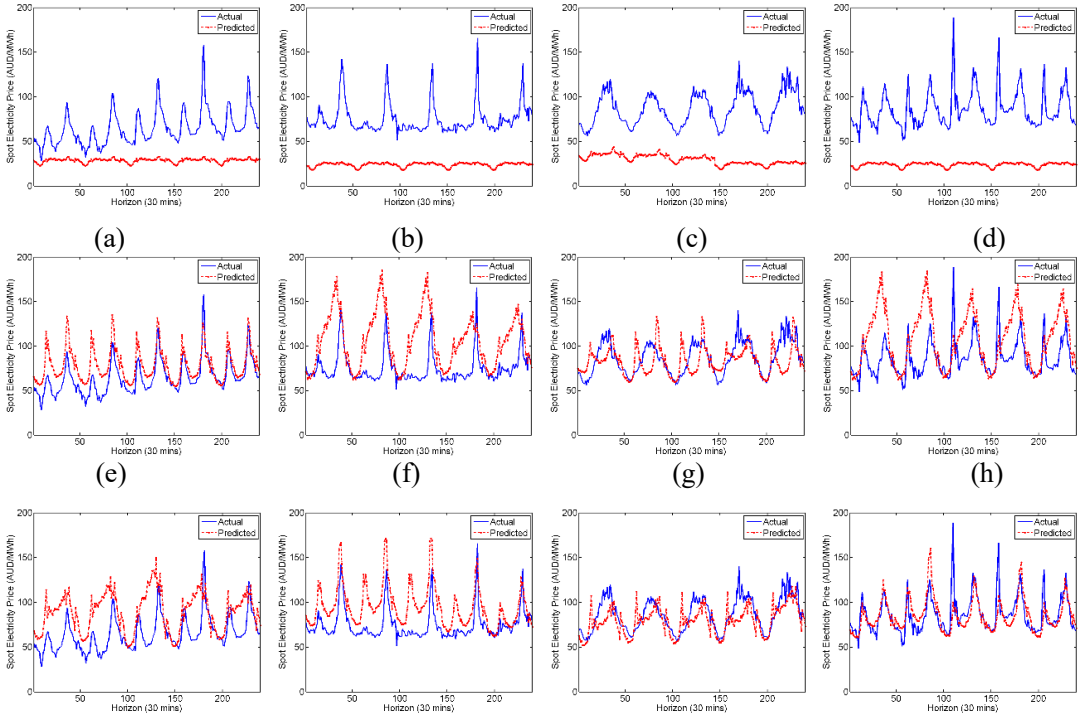
Fig. 6. Results of DM test of forecast models in terms of (a) MAE and (b) RMSE

Fig.6 clearly shows that the proposed CTSGAN is the best forecast model with a small p-value, and BPNN is the worst model. In addition, one can find that the prediction performance of the CTSGAN is statistically significant, with highly saturated green squares in the column labeled CTSGAN. In terms of both the DM test of RMSE, it is not outperformed significantly by the GRU, LASSO, and LSTM.

Fig. A1 shows the autocorrelation error for CTSGAN model, LSTM model, LASSO model and BPNN model, respectively, with a confidence interval of 95%. It can be seen that the autocorrelation is not significant for CTSGAN and LSTM models. The autocorrelation plot shows two spikes at lag 2 and 5 for LASSO, but it is unlikely to have any noticeable impact on forecasts, which is slightly less than 95 percent confidence interval. However, the residuals of BPNN show

large autocorrelation, because of poor forecasting performance.

To compare the models graphically, Fig. 7 plots the rolling forecast (red) and actual (blue) electricity prices for five days in the winter, spring, summer, and autumn seasons for the 2019-2020 financial year by the BPNN, GRU, LSTM, TSGAN, CTSGAN, ARIMA, and LASSO models. As shown in Figs. 7(a)-(d), the BPNN model cannot effectively capture the complex characteristics of electricity prices, and in the four seasons of the financial year 2019-2020, the forecasting results fluctuate between A$20/MWh and A$40/MWh. As shown in Figs. 7(e)-(h), the forecast results of the GRU model are inaccurate for spring and autumn in 2019-2020. The actual autumn daily electricity price usually has two peaks (the blue curve in Fig. 7(h)), while the forecasted result has only one peak (the red curve in Fig. 7(h)). This phenomenon may be attributed to overfitting, where the GRU model learns features that do not belong to each season. The LSTM has also exhibited the same problem of overfitting, e.g., the winter forecasts in Fig. 7(i). The TSGAN model also exhibits poor, unstable forecasting results, as shown in Figs. 7(m)-(p), as it does not include the relevant conditions as the input and relies only on the historical data for forecasting. The forecasting prices given by TSGAN are good for only the days of 28 July 2019 (the first day in Fig. 7(m)), 27 October 2019 (the last day in Fig. 7(n)), 28-29 January 2020 (the first two days in Fig. 7(o)), and 30 April 2020 (the third day in Fig. 7(p)). The ARIMA model provides inaccurate forecasting prices, especially as shown in Fig. 7(v), and the LASSO model forecasts better than the ARIMA model for the winter prices in Fig. 7(y). It is very interesting to note that the GRU, LSTM, ARIMA, and LASSO models all forecast two price peaks incorrectly for the summer prices (Figs. 7(g), (k), (w), (α)), while CTSGAN is able to accurately forecast only one electricity price peak during these five days. The proposed CTSGAN model captures temporal variation very well, with outstanding forecasts of electricity prices for all seasons from 2019-2020 with errors only occurring at the spikes, as shown in Figs. 7(q)-(t).



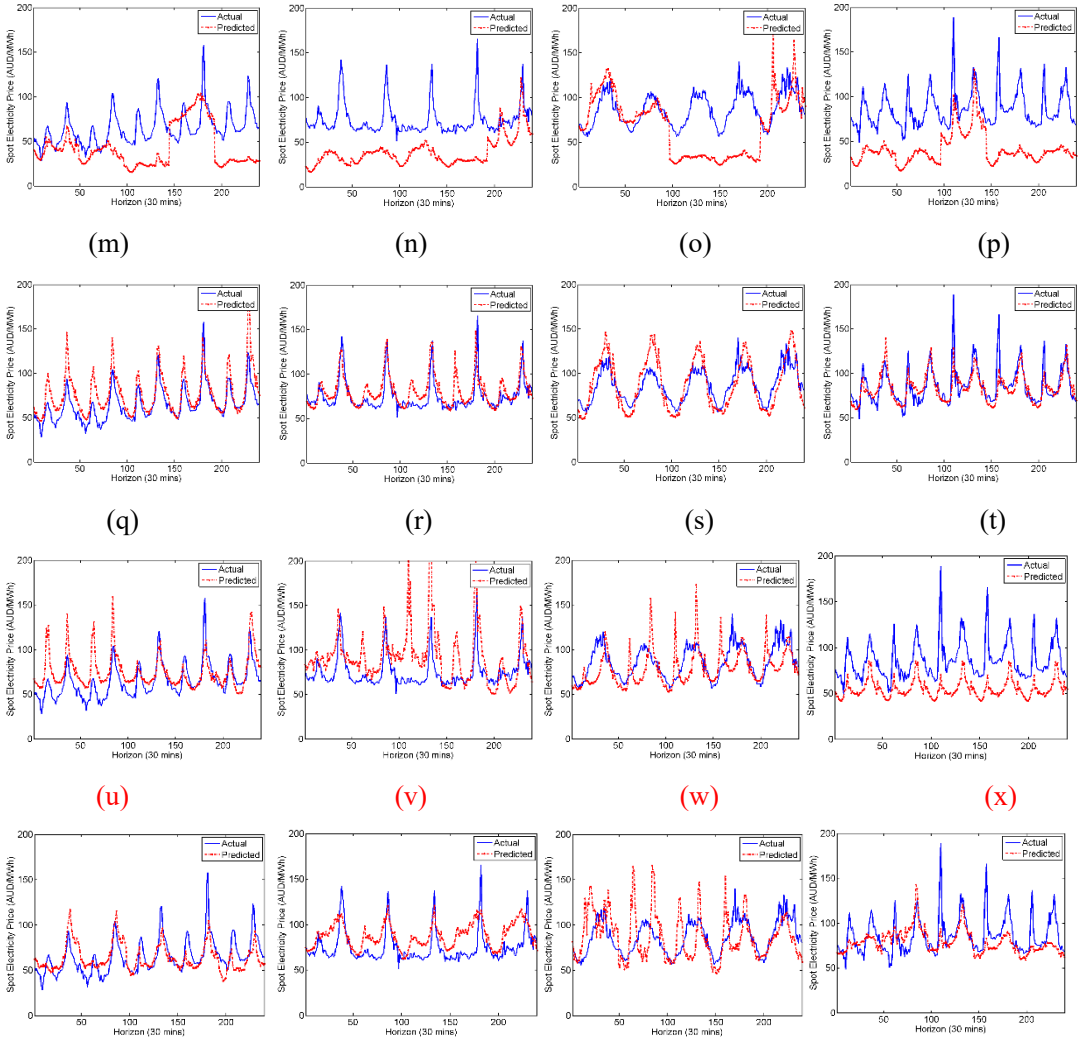


Fig. 7. Day-ahead forecasting for five days in four seasons of the financial year 2019-2020 by (a)-(d) BPNN, (e)-(h) GRU, (i)-(l) LSTM, (m)-(p) TSGAN, (q)-(t) CTSGAN models (u)-(x) ARIMA, and (y)-(β) LASSO

4.1.2 Experiment 2: Verification of generalization ability with CTSGAN

To investigate why the proposed CTSGAN model gives a better prediction than the simple LSTM model, we compare the test and training errors of the CTSGAN and LSTM models for each month in the year 2012. The error gain between the test and training errors, which represents the generalization capability, is calculated.

Table 2. Comparison of CTSGAN and LSTM point forecasting model for each month in 2012

Month	Indicators	CTSGAN			LSTM		
		Test	Training	Gain	Test	Training	Gain
Jan	MAE	**0.0048**	0.0046	1.0445	0.0049	0.0045	1.0968
	MAPE	0.0888	0.0671	1.3229	**0.0875**	0.0668	1.3099
	RMSE	**0.0057**	0.0053	1.0772	0.0059	0.0054	1.0995
	U2	**0.1214**	0.1196	1.0149	0.1232	0.1183	1.0416
	MAE	**0.0064**	0.0061	1.0536	0.0078	0.0060	1.3040

Feb	MAPE	**0.1199**	0.1161	1.0328	0.1418	0.1055	1.3436
	RMSE	**0.0075**	0.0063	1.1933	0.0092	0.0059	1.5569
	U2	**0.1303**	0.1037	1.2567	0.1442	0.1063	1.3568
Mar	MAE	**0.0055**	0.0038	1.4444	0.0084	0.0039	2.1559
	MAPE	**0.0929**	0.0637	1.4592	0.1390	0.0529	2.6275
	RMSE	**0.0064**	0.0048	1.3391	0.0097	0.0043	2.2619
	U2	**0.1223**	0.0998	1.2255	0.1793	0.0947	1.8930
Apr	MAE	**0.0053**	0.0052	1.0250	0.0061	0.0051	1.1942
	MAPE	**0.1022**	0.1004	1.0174	0.1130	0.0831	1.3600
	RMSE	**0.0062**	0.0049	1.2663	0.0072	0.0047	1.5219
	U2	**0.1221**	0.1104	1.1061	0.1395	0.1122	1.2430
May	MAE	**0.0111**	0.0083	1.3407	0.0156	0.0078	2.0024
	MAPE	**0.1449**	0.1271	1.1404	0.2016	0.1284	1.5700
	RMSE	**0.0128**	0.0103	1.2386	0.0171	0.0093	1.8420
	U2	**0.1534**	0.1498	1.0239	0.1922	0.1577	1.2187
Jun	MAE	**0.0068**	0.0038	1.7894	0.0082	0.0044	1.8743
	MAPE	**0.1085**	0.0841	1.2900	0.1482	0.0749	1.9783
	RMSE	**0.0085**	0.0064	1.3243	0.0096	0.0057	1.6913
	U2	**0.1695**	0.0931	1.8202	0.2342	0.0893	2.6224
Jul	MAE	0.0219	0.0114	1.9232	**0.0111**	0.0108	1.0278
	MAPE	0.2371	0.1113	2.1305	**0.1288**	0.1136	1.1336
	RMSE	0.0261	0.0134	1.9492	**0.0150**	0.0133	1.1302
	U2	0.1660	0.0978	1.6974	**0.1068**	0.0922	1.1587
Aug	MAE	**0.0163**	0.0143	1.1380	0.0173	0.0144	1.1989
	MAPE	**0.1413**	0.1294	1.0920	0.1521	0.1309	1.1619
	RMSE	**0.0222**	0.0221	1.0068	0.0229	0.0222	1.0299
	U2	0.1052	0.1039	1.0126	**0.1063**	0.1027	1.0350
Sep	MAE	**0.0105**	0.0103	1.0210	0.0108	0.0101	1.0735
	MAPE	**0.1291**	0.1082	1.1929	0.1353	0.0922	1.4676
	RMSE	**0.0131**	0.0116	1.1267	0.0140	0.0121	1.1542
	U2	**0.1847**	0.1655	1.1161	0.2074	0.1672	1.2404
Oct	MAE	0.0078	0.0078	0.9993	**0.0073**	0.0083	0.8810
	MAPE	**0.1104**	0.1021	1.0814	0.1048	0.1013	1.0349
	RMSE	**0.0094**	0.0063	1.4950	0.0095	0.0064	1.4795
	U2	**0.1810**	0.1733	1.0445	0.1987	0.1536	1.2936
Nov	MAE	**0.0145**	0.0113	1.2836	0.0270	0.0125	2.1624
	MAPE	**0.1638**	0.1333	1.2289	0.3221	0.2001	1.6097
	RMSE	**0.0222**	0.0208	1.0655	0.0313	0.0209	1.4976
	U2	**0.1749**	0.1677	1.0429	0.2804	0.1631	1.7191
Dec	MAE	**0.0170**	0.0143	1.1846	0.0249	0.0154	1.6206
	MAPE	**0.1713**	0.1093	1.5670	0.3135	0.0978	3.2058
	RMSE	**0.0241**	0.0198	1.2170	0.0286	0.0152	1.8806
	U2	**0.3100**	0.2931	1.0577	0.5681	0.2992	1.8986
Average	MAE	**0.0107 (14.4%)**	**0.0084**	**1.2706**	0.0125	0.0086	1.4660
	MAPE	**0.1342 (19.0%)**	0.1043	**1.2963**	0.1656	**0.1040**	1.6502
	RMSE	**0.0137 (8.7%)**	0.0110	**1.2749**	0.0150	**0.0104**	1.5121
	U2	**0.1617 (21.7%)**	0.1398	**1.2015**	0.2067	**0.1380**	1.4768

Table 2 shows a monthly analysis of the forecasted electricity prices in 2012. The values in brackets in the bottom four cells of the CTSGAN/Test Column represent the improvement of error by using the CTSGAN model compared to the LSTM model, elucidating that the proposed CTSGAN model can give overall better forecasts for the test set. However, the forecasts in July are worse than those given by the LSTM model. The errors of forecasts given by the CTSGAN model

are 0.0219 (MAE), 0.2371 (MAPE), 0.0261 (RMSE) and 0.1660 (U2), which are larger than those of the simple LSTM model by 0.0111 (MAE), 0.1288 (MAPE), 0.0150 (RMSE) and 0.1068 (U2), respectively. The reason for this may be attributed to the introduction of carbon pricing in July. Although carbon pricing has been removed from our model, it still has a profound effect not only on prices but also on the amount of electricity generation and demand. In July, the first month of introducing the carbon tax, the generator in the proposed CTSGAN fits not well, and the discriminator requires more training to effectively distinguish between the real data and the generated data under the data mutation. Also, it can be found that the difference in fitting error with two models for the July training set is not significant, and only the generalization ability is affected.

After calculating the average of the training and test errors for each of the 12 months, it is evident that the proposed CTSGAN model can forecast better for the test set, with the average MAE of 0.0107, MAPE of 0.1342, RMSE of 0.0137, and U2 of 0.1617. Compared to the simple LSTM model, the 12-month average MAE, MAPE, RMSE, and U2 are improved by 14.4%, 19.0%, 8.7%, and 21.7%, respectively. However, when comparing the fitting errors of the training set, the fitting errors of the LSTM model are smaller, with the average MAPE of 0.1040, the average RMSE of 0.0104, and the average U2 of 0.1380. Overall, the LSTM model is an excellent fitting tool for the time series due to the selection of long and short network memory ability, resulting in a minimal training error, but at the same time may introduce the problem of overfitting. For the proposed CTSGAN model, due to the latent space, there are two other network errors, the embedding error and the recovery error. Compared to the simple LSTM model, the overall fitting error is inferior to that of the LSTM model.

Furthermore, the CTSGAN/Gain Column and LSTM/Gain Column, obtained by dividing the test error by the training error, represent the generalization ability based one the CTSGAN model and the LSTM model. The smaller the value, the better the generalization ability. Fig. 8 illustrates the gains with CTSGAN and LSTM in terms of MAE, MAPE, RMSE, and U2 in the year 2012.

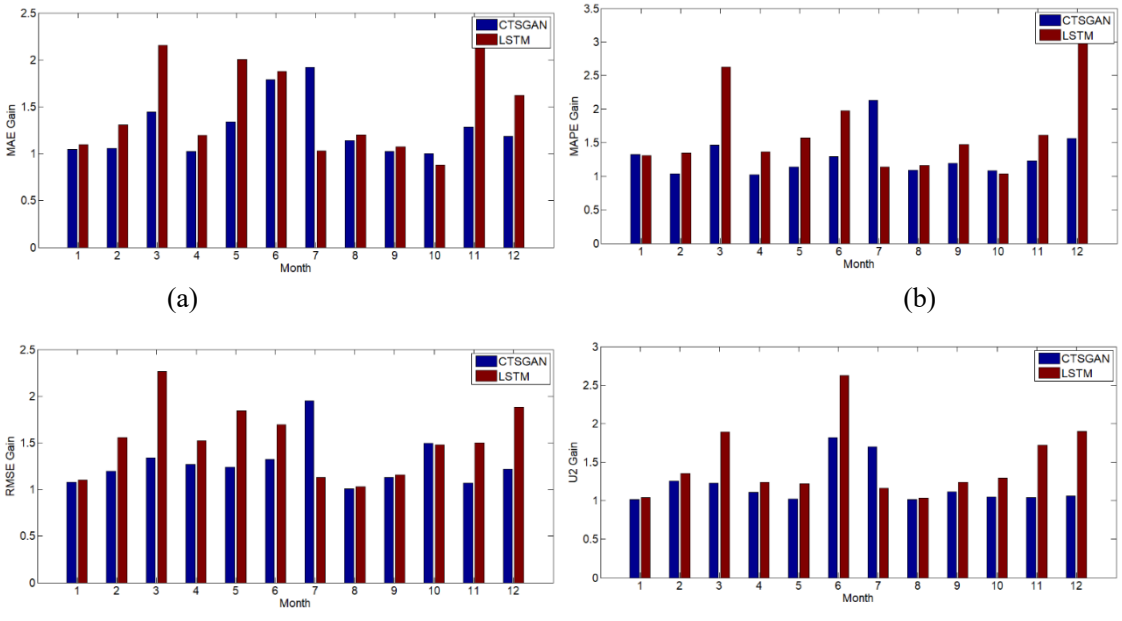


Fig. 8. Monthly gains between CTSGAN and LSTM in terms of
(a) MAE (b) MAPE (c) RMSE (d) U2

From Figs. 8(b), compared to the gain between testing MAPE error and training error of LSTM, the gain of CTSGAN is significantly smaller, for example, which is only half of that of LSTM, with a MAPE of 1.45 and 1.57 in March and December, respectively. In terms of MAE, RMSE, and U2, the gains based on CTSGAN are much smaller in most months of the whole year, which means that the CTSGAN model has a better generalization ability than the LSTM model. Although the training error of the CTSGAN model is not the same as that of the LSTM model, the gain of the CTSGAN model is significantly better than that of the LSTM model due to the generalization ability provided by the discriminator. Since the latent space retains the temporal dynamics, which is an excellent improvement to the forecasting ability, the point forecasting based on the CTSGAN model is effective.

4.1.3 Experiment 3: Verification of optimal condition selection method and effect of removal of spikes

In Section 3.2.2, we propose a PSO-LSTM condition selection method to choose the optimal condition set as the input of CTSGAN due to the core NNs of CTSGAN are LSTMs. In this section, the forecasts of electricity prices for the year 2019 are conducted with historical electricity price set, all condition set, and the optimal condition set, respectively, to verify the effectiveness of the optimal condition selection method. Fig. 9 plots the bar charts of forecasting errors with three different condition sets, and Table 3 lists the results of four indicators and improvements.

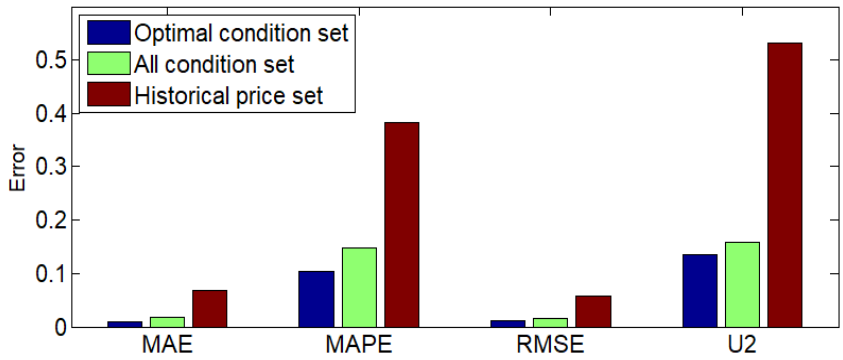


Fig. 9. Forecasting errors with optimal condition set and other two different condition sets

In Fig. 9, one can find that when the CTSGAN input condition set only contains historical prices, the forecasting errors almost double or triple the errors with all condition sets as input. However, all condition set contains too many irrelevant variables, and DL will learn the features and result in poorer forecasting results. The optimal condition set selected by PSO, which is applied as the input of the CTSGAN model, provides the best forecasting performance.

Table 3. Comparison of different condition sets as input

Year	Indicators	Optimal condition set	All condition set	Historical price set
2019	MAE	**0.0093 (47.7%, 86.4%)**	0.0178	0.0682
	MAPE	**0.1034 (29.8%, 73.1%)**	0.1473	0.3839
	RMSE	**0.0108 (32.1%, 81.2%)**	0.0159	0.0574
	U2	**0.1362 (13.9%, 74.4%)**	0.1581	0.5312

The values in brackets in cells of the optimal condition set column represent the improvement of error by introducing the optimal condition set compared to all condition set and historical price set, elucidating that the PSO-LSTM condition selection method can reduce forecasting errors. Compared with all condition set input, MAE, MAPE, RMSE, and U2 are improved by 47.7%,

29.8%, 32.1%, and 13.9%, respectively. When the condition input is only historical price set, the CTSGAN model provides the worst prediction error, with MAE of 0.0682, MAPE of 0.3839, RMSE of 0.0774, and U2 of 0.5312, all of which are several times larger than those with optimal condition set.

Table 4. Comparison of effect of removing spikes on forecasts

Year	Indicators	w/o positive w/o negative	w/ positive w/o negative	w/o positive w/ negative	w/ positive w/ negative
2019	MAE	**0.0093 (92.6%, 32.6%, 93.3%)**	0.1248	0.0138	0.1392
	MAPE	**0.1034 (76.6%, 16.5%, 78.9%)**	0.4422	0.1239	0.4904
	RMSE	**0.0108 (87.8%, 13.6%, 89.5%)**	0.0891	0.0125	0.1025
	U2	**0.1362 (82.5%, 29.4%, 84.7%)**	0.7803	0.1930	0.8917

Table 4 lists the effect of the removal of positive and negative spikes on forecasts. The values in brackets in cells of the "w/o positive w/o negative" column represent the improvement of error by removing all spikes compared to removing one or none spikes. One can find that the removal of positive and negative spikes can improve forecasting performance. Compared with the removal of none spikes data (original prices), MAE, MAPE, RMSE, and U2 are improved by 93.3%, 78.9%, 89.5%, and 84.7%, respectively. Another finding is that removing only positive spikes resulted in a greater boost than removing only negative spikes, with a RMSE of 0.0891 (w/o negative) and 0.0125 (w/o positive). The main reason is that the positive spikes are large (>10000A$/MWh) and if not removed will result in the normalized electricity rices becoming very small, making it difficult to fit the prediction model. Negative electricity price spikes, on the other hand, are small (>-100A$/MWh) and have limited effect on forecasts.

4.1.4 Experiment 4: Verification of contribution of each component in CTSGAN

Table 5. Contribution of each component in CTSGAN

	2012	2014	2016	2018
CTSGAN	0.0137±0.0061	0.0132±0.0043	0.0456±0.0074	0.0372±0.0036
w/o conditions	0.0539±0.0073	0.0574±0.0047	0.0749±0.0094	0.0974±0.0055
w/o Wasserstein	0.0377±0.0143	0.0465±0.0122	0.0691±0.0101	0.0571±0.0098
w/o supervised loss	0.0471±0.0078	0.0493±0.0068	0.0731±0.0073	0.0837±0.0061
w/o embedding networks	0.0289±0.0086	0.0316±0.0094	0.0592±0.0082	0.0629±0.0063
w/o joint training	0.0363±0.0116	0.0485±0.0127	0.0605±0.0099	0.0774±0.0122

To analyze the contribution of each component in CTSGAN, RMSE is used as a point prediction benchmark with the following modifications: (1) without conditions; (2) without Wasserstein Distance mechanism; (3) without the supervised loss; (4) without the embedding networks; and (5) without joint training process. Table 5 lists the RMSE results of the ablation experiment. Overall, one can find that all five elements make important contributions in improving the quality of electricity prices forecasting. The condition set plays one of the most important roles in improving the accuracy of the prediction results. Taking the 2014 price prediction results as an example, one can find that the ranking according to the error range contribution is in the order of the joint training process (±0.0127), Wasserstein Distance (±0.0122), embedding networks (±0.0094), and supervised loss (±0.0068), which well reveals the reason for the stability of the proposed CTSGAN.

4.2 Case 2: Electricity price probabilistic forecasting

To verify the probabilistic forecasting performance of the proposed CTSGAN model, the TSGAN model is introduced for forecasting and comparison in Section 4.2.1 and 4.2.2. The parameter setting and selection of core NNs are the same as described in Section 4.1. Different from point forecasting, probabilistic forecasting chose 100 as the noise dimensionality. In Section 4.2.3, Bootstrap and LUBE-based models are introduced to compared with CTSGAN model for winter and summer electricity price forecasting.

4.2.1 Experiment 5: Comparison of synthetic scenarios with CTSGAN and TSGAN

Fig. 10 illustrates the PCA and t-SNE visualization on the synthetic scenarios (blue) and original (red) electricity prices with (a) and (b) the proposed CTSGAN model, and (c) and (d) the TSGAN model. The scenarios generated by the proposed CTSGAN show a better overlap with the original electricity price than the TSGAN model. Comparing Figs. 10(a) and (c), one can find that the original data is clustered into two concentrated regions, and the generated data from the CTSGAN model can also be clustered into two regions, while the generated data from the TSGAN model does not have this feature. Besides, the data generated by the CTSGAN model has a good overlap of features outside of the aggregated regions, while the data from the TSGAN model has lost this feature. The same conclusion can be found for Figs. 10(b) and (d), where the data generated by the CTSGAN model can retain more original features and overlap better with the original electricity price data than those by TSGAN. The CTSGAN model can make the generated price scenarios relatively realistic.

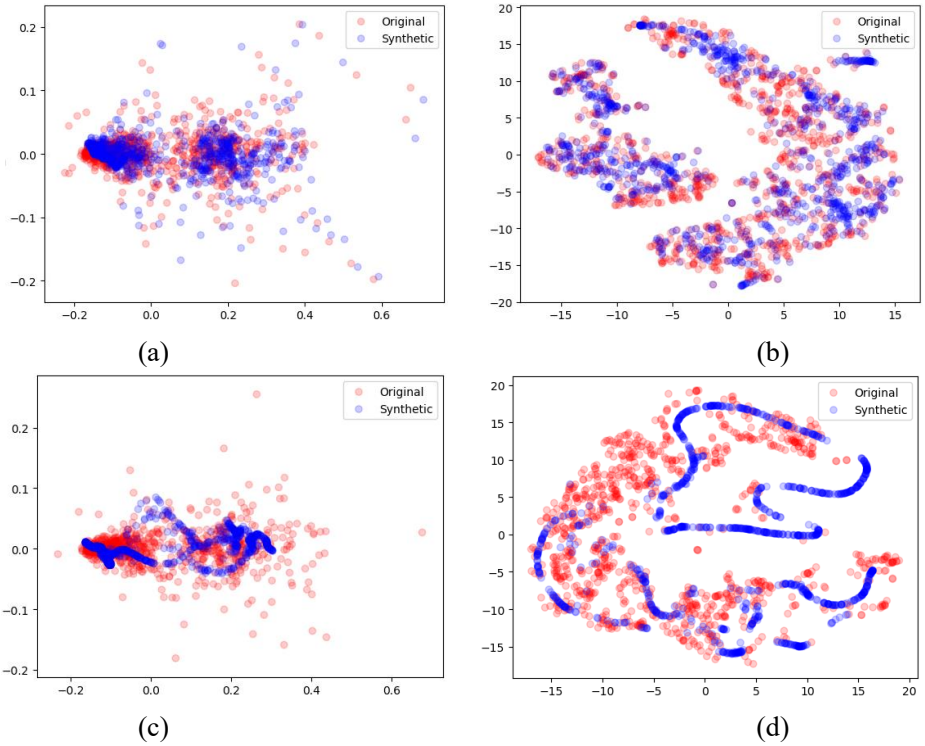


Fig.10. Visualization with t-SNE and PCA for synthetic scenarios based on (a) and (b) CTSGAN, and (c) and (d) TSGAN

4.2.2 Experiment 6: Comparison of probabilistic forecasting with CTSGAN and TSGAN

The high and fluctuating electricity prices during the autumn and winter months due to the high

heating demand makes it difficult to forecast. Therefore, the autumn and winter seasons are chosen for a probabilistic forecasting case study. Figs. 11 and 12 show forecasts of 1,000 scenarios of 5 days for the winter and autumn seasons in the financial years of 2018-2019 and 2019-2020, respectively.

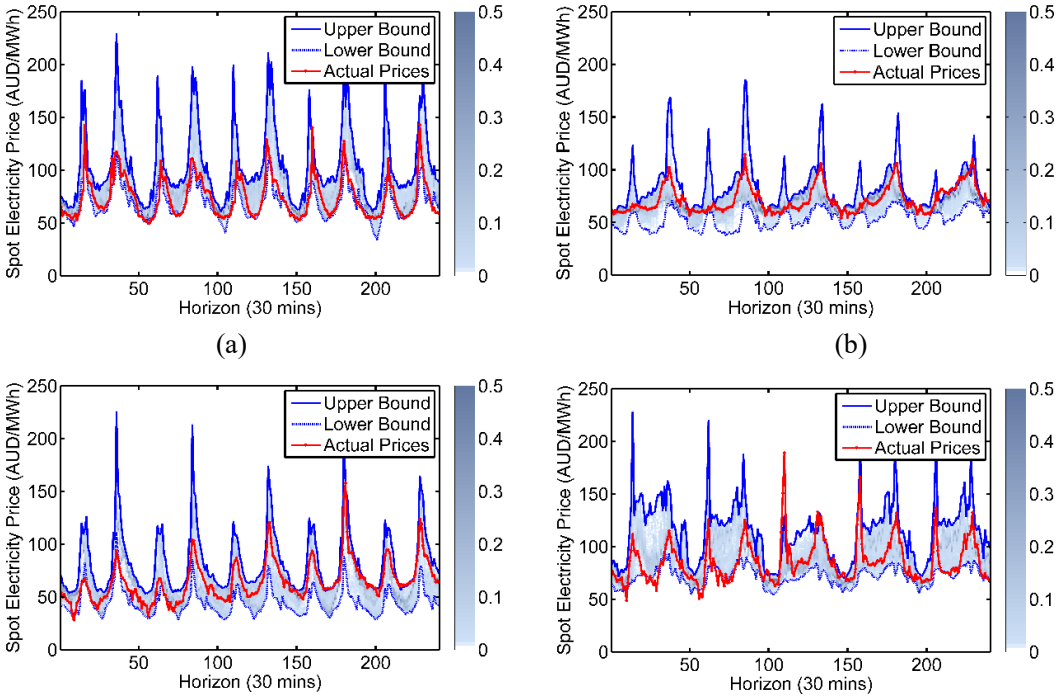


Fig. 11. Day-ahead probabilistic forecasting for five days based on CTSGAN, (a) Winter in 2018-2019, (b) Autumn in 2018-2019, (c) Winter in 2019-2020, (d) Autumn in 2019-2020

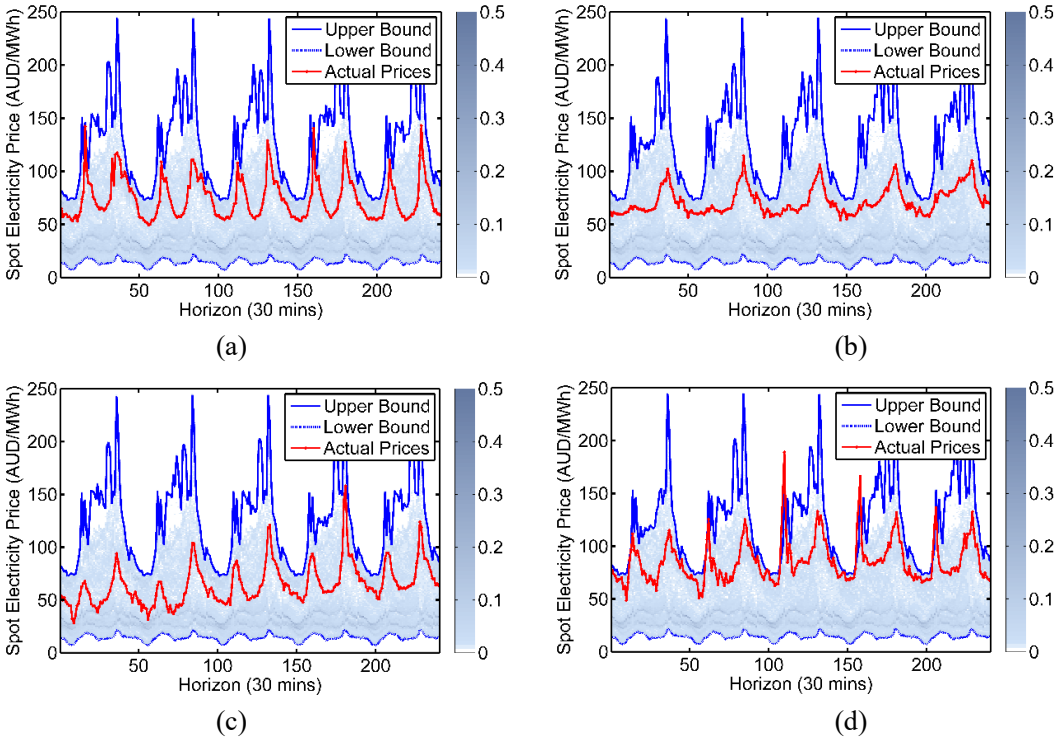


Fig. 12. Day-ahead probabilistic forecasting for five days based on TSGAN, (a) Winter in 2018-2019, (b) Autumn in 2018-2019, (c) Winter in 2019-2020, (d) Autumn in 2019-2020

The blue solid and dotted curves represent the upper and lower bounds of the 1,000 scenarios, respectively. The probability of each electricity price occurring is calculated with 1,000 scenarios, and the predicted probability of each occurrence is represented with a blue-grey pixel. As shown in the color bar (the legend on the right), the darker the color, the higher the probability. As there are few points with a probability greater than 0.5, the range of the color bar is set as 0-0.5, and the white pixel is shown for the probability of 0.

As shown in Figs. 11 and 12, the variances of forecasting scenarios are significantly more extensive at the price peaks. For example, in Fig. 11(a), the maximum value of each spike forecast is close to A$200/MWh, while the minimum value of that is only around A$100/MWh. However, in the low-price areas, this fluctuation is smaller, within a range between A$50/MWh and A$70/MWh, as shown in Fig. 11(a). The same phenomena can be seen in Fig. 12. Another noticeable phenomenon is that the forecast intervals in Fig. 12 are significantly larger than those in Fig. 11, with the interval widths even more remarkable than A$200/MWh at the price peak and greater than A$50/MWh in the low price areas. While a large interval maximizes the inclusion of actual electricity prices, the forecast accuracy is significantly reduced. A A$200/MWh width interval forecast is not meaningful in many cases. The high probability pixel in Fig. 11 is more concentrated around the actual electricity prices. This phenomenon is even more evident in Fig. 11(b), where the large probability forecast pixels almost coincide with the actual electricity prices, from about the 144-th point on the horizontal axis to the 240-th point (1 May 2019-2 May 2019). Conversely, the large probability forecast pixels in Fig. 12 are mainly concentrated below 50, deviating from the actual electricity prices and challenging the forecast effectively.

Table 6 lists the performances of the proposed CTSGAN model and the TSGAN model for probabilistic forecasting of day-ahead electricity prices in terms of CRPS, WS with α=0.1, and WS with α=0.2. As shown, the CTSGAN model outperforms the TSGAN model, with an average CRPS of 7.0371, WS of 10.6850 (α=0.1), and WS of 10.0204 (α=0.2).

Table 6. Comparison of CTSGAN and TSGAN probabilistic forecasting model for each season

Year	Season	Model	CRPS	WS α=0.1	WS α=0.2
2018	Winter	CTSGAN	**6.8006**	**9.0710**	**8.7695**
-		TSGAN	18.2032	57.0985	48.9610
2019	Spring	CTSGAN	**7.8141**	**8.9321**	**5.5526**
		TSGAN	16.8523	57.3740	54.5402
	Summer	CTSGAN	**3.4887**	**9.4051**	**9.3615**
		TSGAN	17.5536	58.9055	52.4945
	Autumn	CTSGAN	**6.0216**	**10.2391**	**9.9875**
		TSGAN	18.6793	59.0754	55.7626
2019	Winter	CTSGAN	**9.2235**	**11.6335**	**11.1395**
-		TSGAN	18.8676	58.2664	57.1147
2020	Spring	CTSGAN	**8.3984**	**11.4421**	**11.1248**
		TSGAN	22.5368	59.0245	57.1315
	Summer	CTSGAN	**7.0288**	**13.7510**	**13.5858**
		TSGAN	22.4623	57.7852	55.2205
	Autumn	CTSGAN	**7.5209**	**11.0063**	**10.6418**

	TSGAN	16.1978	58.3173	49.4602
Average	CTSGAN	**7.0371**	**10.6850**	**10.0204**
	TSGAN	18.9191	58.2309	53.8356

Overall, the price scenarios generated by the proposed CTSGAN model are more concentrated compared with the TSGAN model. The main reason is that the introduction of conditions improves the forecasting accuracy, and the latent space further preserves the temporal dynamics. If the difference between the forecasted scenarios and actual prices is too significant, the discriminator can readily recognize the forecasted scenarios.

4.2.3 Experiment 7: Comparison of probabilistic forecasting with different models

Conventional probabilistic forecasting models like Bootstrap, LUBE, QR and KDE introduced to benchmark the proposed CTSGAN model. The parameters of ELM combined Bootstrap (ELM-Bootstrap) are set as [64], and three LUBE models are constructed by ANN (ANN-LUBE) [65], WNN (WNN-LUBE) [13], RNN (RNN-LUBE) [26], respectively. The parameters of QRRF and KDE are set as [66] and [67]. Table 7 lists the performances of five different models for probabilistic forecasting in terms of CRPS, WS with α=0.1, and WS with α=0.2.

Table 7. Comparison of four probabilistic forecasting models for winter and summer sets

Year	Seasons	Model	CRPS	WS α=0.1	WS α=0.2
2018	Winter	CTSGAN	**6.8006**	**9.0710**	**8.7695**
-		ELM-Bootstrap	17.8312	48.9931	40.7897
2019		ANN-LUBE	13.9349	25.8316	23.9824
		WNN-LUBE	10.7461	20.9943	19.1923
		RNN-LUBE	10.3832	18.9492	17.5467
		QRRF	10.8467	16.7312	14.2791
		QRRF-KDE	8.9827	13.3732	12.6639
	Summer	CTSGAN	**3.4887**	**9.4051**	**9.3615**
		ELM-Bootstrap	14.8311	36.1319	34.0389
		ANN-LUBE	10.3785	30.4334	29.7812
		WNN-LUBE	10.8423	31.0327	28.9314
		RNN-LUBE	8.7303	30.1924	28.8902
		QRRF	8.0371	26.7723	21.8369
		QRRF-KDE	7.7990	23.9754	20.6363
2019	Winter	CTSGAN	**9.2235**	**11.6335**	**11.1395**
-		ELM-Bootstrap	19.0395	59.8314	55.8831
2020		ANN-LUBE	14.7873	25.3746	23.2423
		WNN-LUBE	14.5523	22.2537	20.3784
		RNN-LUBE	13.3936	22.8848	20.6642
		QRRF	12.1114	21.6312	20.1977
		QRRF-KDE	11.6700	19.9081	19.0834
	Summer	CTSGAN	**7.0288**	**13.7510**	**13.5858**
		ELM-Bootstrap	19.7371	47.8456	44.9173
		ANN-LUBE	15.3314	40.6649	38.8943
		WNN-LUBE	10.8731	29.6764	24.5238

RNN-LUBE	10.1731	25.8837	21.4831
QRRF	8.2726	22.3005	16.8893
QRRF-KDE	7.9003	21.7733	15.1023

The results show that the proposed CTSGAN model has the best performance in the electricity price interval forecasting for winter and summer from 2018 to 2019 and 2019 to 2020, giving a smaller error than other models. For example, the WS (α=0.1) is 9.0710 (winter, 2018-2019), 9.4051 (summer, 2018 -2019), 11.6335 (winter, 2019-2020), and 13.7510 (summer, 2019-2020), while that obtained by Bootstrap models and LUBE models are approximately two times or three times than that obtained by proposed CTSGAN model. The QRRF predictor, with a CRPS of 8.2726, outperforms all LUBE models, but is slightly inferior to QRRF-KDE, with a CRPS of 7.9003, in the summer from 2019 to 2020. Compared forecasting accuracy in summer with that in winter, one can find that the proposed CTSGAN model provides a smaller CRPS in summer, which means that the forecasting intervals in summer have higher reliability and lower sharpness.

4.2.4 Experiment 8: Probability density of prediction intervals based on CTSGAN

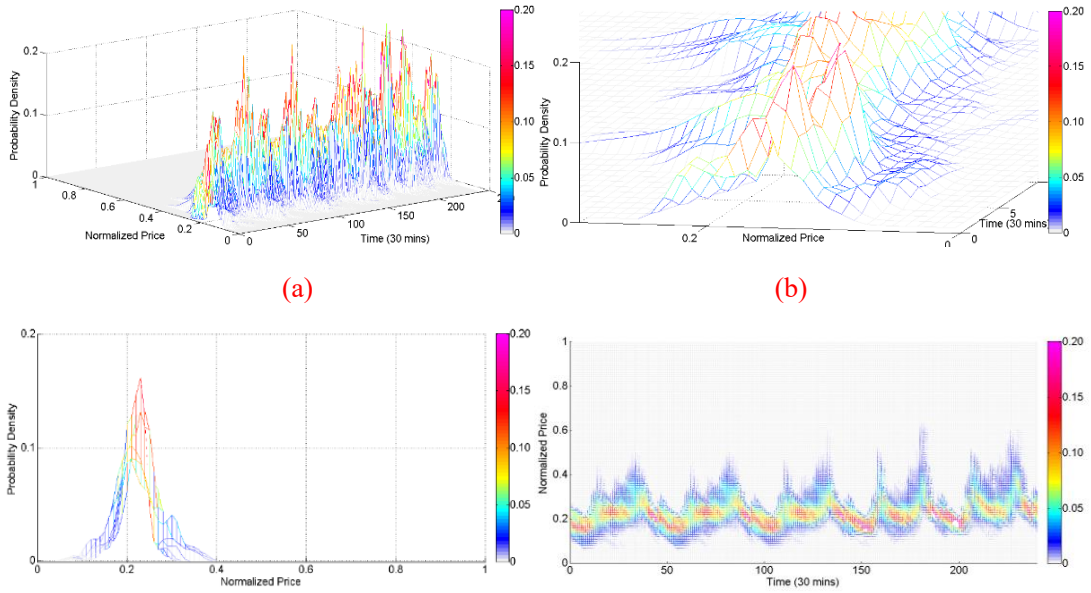


Fig. 13. Probability density of prediction scenarios
(a) Overall view,(b) Enlarged partial view, (c) Left view (d) Top view

The conventional interval prediction methods, such as the LUBE-based methods, only give the upper and lower bounds of prediction intervals but cannot give the probability density of each electricity price in the interval. Our proposed CTSGAN method can generate prediction intervals and probability densities by combining the pattern-diversity scenarios. The visualized prediction scenarios are shown in Fig. 13(a), where the x, y, and z axes are the prediction time scale, the normalized prediction electricity price, and the probability density for each different price, respectively. The colors represent the probability densities of the occurrence of these scenarios. The x-axis [0,10] in Fig. 13(a) is partially enlarged, as shown in Fig. 13(b), which presents the detailed probability density of each scenario. The probabilities vary from time to time for different scenarios, with the highest probabilities ranging from 0.15 to 0.2. Fig. 13(c) presents the left view of Fig. 13(b). As shown, the probability density cross-section of each scenario for different times approximates a

Gaussian distribution curve. This indicates that the pattern-diversity scenarios generated by CTSGAN follow approximately the normal distribution, consistent with the input random noise distribution. Fig. 13(d) is a top view of Fig. 13(a). Fig. 13(d) shows how scenarios with different probability densities can be combined into prediction intervals. Similar to Fig. 13(c), the scenarios with greater probabilities appear at the center of the prediction intervals, while scenarios of low probabilities stay near the upper and lower boundaries.

4.2.5 Experiment 9: Simulations of multi-objective forecasting with different dimensional random noise

By varying the random noise dimensionality of the input, the proposed CTSGAN model can be transformed into a multi-objective forecasting model, which directly gives high-quality forecasting intervals with different coverage probabilities. The forecast of electricity prices is repeated 11 times for four seasons in the financial years of 2018-2019 and 2019-2020, setting the dimensionality of random noise to 40, 60, 80, 100, 120, 140, 160, 180, 200, 220 and 240, respectively. The interval widths of the 1,000 generated scenarios are then calculated and compared with the actual electricity prices to determine whether the forecasting interval covers the actual prices and calculate the coverage probability. The result is shown in Fig. 14.

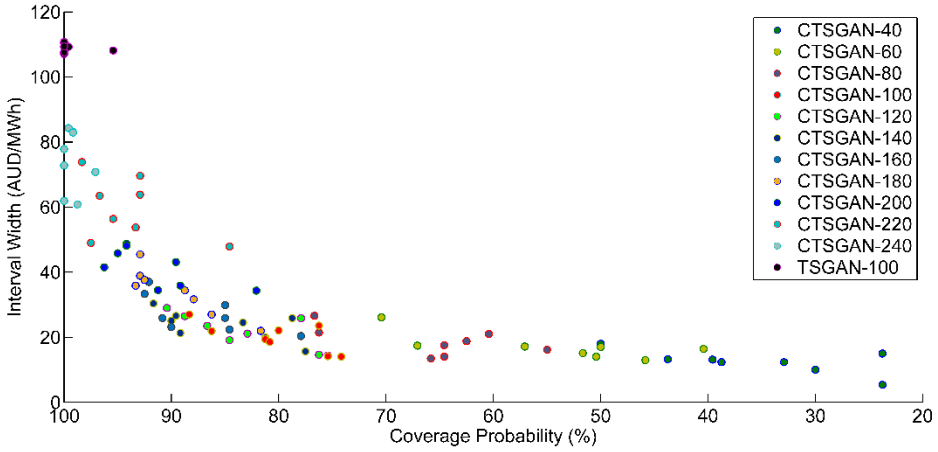


Fig. 14. Probabilistic forecasting with different dimensional random noise

In Fig. 14, the different colored points represent different random noise input dimensions, with 240 on the far left and 40 on the far right. There are 8 points for each color, representing the results of four seasons in the two financial years of 2018-2019 and 2019-2020. The multiple red points are the results obtained with 100-dimensional noise as the input, and the green points are the results obtained with the 120-dimensional noise. Thus, taking 100 or 120 as the input random noise dimension can guarantee a higher coverage probability without reducing the accuracy (wider intervals). In this paper, 100 is chosen. When the dimension of input random noise increases, the increased diverse randomness would diversify the forecasted scenarios, and therefore, the forecast interval becomes wider, with the interval width essentially greater than A$60/MWh when the dimensionality is 240. As the diversity of the generated scenarios gradually covers some small probability prices, the coverage probability increases, with a probability of about 95% when the dimensionality is 240. Conversely, as the dimensionality of the input random noise decreases, the diversity of the generated scenarios also decreases. The forecasted interval becomes progressively

narrower, and the probability of coverage decreases, with a coverage probability of less than 60% and a forecasted interval width of less than A$20/MWh when the dimensionality is 40 or 60.

The black points with the red outlines in the top left corner of Fig. 14 are the forecasted interval width and probability results based on the TSGAN model, with 100-dimensional noise as the input. This result is relatively poor. Although the coverage probability is large, the width of the forecasted interval is also large, equal to about A$115/MWh, even greater than the interval width of the output based on the CTSGAN model with 240 dimensions of noise as the input (about A$60/MWh to A$85/MWh). Consequently, the proposed CTSGAN model is more appropriate for obtaining different forecasting results by changing the input dimensions, which is not the case with the TSGAN model.

5. Conclusion and future work

Accurate electricity price forecasting plays an essential role in the electricity market. In this paper, a new CTSGAN-based electricity price forecasting model is proposed, which can well preserve the temporal dynamics for the whole period. This model exhibits the excellent performances of both point and probabilistic forecasting according to case studies.

Firstly, for point forecasting, the proposed CTSGAN model has better forecasting accuracy than other conventional ML or DL models and has better generalization ability than the simple LSTM model due to the introduction of the discriminator. Secondly, compared with different condition sets as input, the effectiveness of PSO based optimal condition selection method is verified. Thirdly, for probabilistic forecasting, the CTSGAN model can generate relatively realistic price scenarios, which retains more of the original features and has a better overlap with the original electricity prices. And the CTSGAN model has better probabilistic forecasting performance than the TSGAN model in terms of CRPS and WS. Furthermore, other conventional probabilistic forecasting models are introduced to compare with the CTSGAN model, validating the effectiveness. Finally, by varying the random noise dimensionality of the input, the proposed CTSGAN model can directly yield high-quality forecasting intervals with different coverage probabilities.

Future works will be conducted on the following two aspects. The first is the improvement and optimization of the forecasting model. At present, deep learning is getting progressively larger, and larger-scale prediction models will undoubtedly improve the accuracy of forecasts. The second is the inclusion of some factors that were not previously included but can significantly affect the electricity prices, such as Covid-19 and the 2019-2020 NSW bushfires.

References

[1] An J, Lee M, Yeom S, Hong T. Determining the Peer-to-Peer electricity trading price and strategy for energy prosumers and consumers within a microgrid. Applied energy. 2020;261:114335.

[2] Bagchi A, Goel L, Wang P. Adequacy Assessment of Generating Systems Incorporating Storage Integrated Virtual Power Plants. IEEE transactions on smart grid. 2019;10:3440-51.

[3] Liang Z, Alsafasfeh Q, Jin T, Pourbabak H, Su W. Risk-Constrained Optimal Energy Management for Virtual Power Plants Considering Correlated Demand Response. IEEE transactions on smart grid. 2019;10:1577-87.

[4] Heydari A, Nezhad MM, Pirshayan E, Garcia DA, Keynia F, De Santoli L. Short-term electricity price and load forecasting in isolated power grids based on composite neural network and gravitational search optimization algorithm. Applied Energy. 2020;277:115503.

[5] Zhang J, Tan Z, Wei Y. An adaptive hybrid model for short term electricity price forecasting. Applied Energy. 2020;258:114087.
[6] Corradi O, Ochsenfeld H, Madsen H, Pinson P. Controlling Electricity Consumption by Forecasting its Response to Varying Prices. IEEE transactions on power systems. 2013;28:421-9.
[7] Shao Z, Zheng Q, Yang S, Gao F, Cheng M, Zhang Q, et al. Modeling and forecasting the electricity clearing price: A novel BELM based pattern classification framework and a comparative analytic study on multi-layer BELM and LSTM. Energy economics. 2020;86:104648.
[8] Uniejewski B, Weron R, Ziel F. Variance Stabilizing Transformations for Electricity Spot Price Forecasting. IEEE transactions on power systems. 2018;33:2219-29.
[9] Chang Z, Zhang Y, Chen W. Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. Energy (Oxford). 2019;187:115804.
[10] Kempitiya T, Sierla S, De Silva D, Yli-Ojanperä M, Alahakoon D, Vyatkin V. An Artificial Intelligence framework for bidding optimization with uncertainty in multiple frequency reserve markets. Applied energy. 2020;280.
[11] Dong Y, Wang J, Jiang H, Wu J. Short-term electricity price forecast based on the improved hybrid model. Energy conversion and management. 2011;52:2987-95.
[12] Uniejewski B, Marcjasz G, Weron R. On the importance of the long-term seasonal component in day-ahead electricity price forecasting; Part II -- Probabilistic forecasting. Energy economics. 2019;79:171.
[13] Chen J, Shen Y, Lu X, Ji Z. An intelligent multi-objective optimized method for wind power prediction intervals. Power System Technology. 2016;40:2758-65.
[14] Yanxia S, Xu W, Jie C. Wind Power Forecasting Using Multi-Objective Evolutionary Algorithms for Wavelet Neural Network-Optimized Prediction Intervals. Applied sciences. 2018;8:185.
[15] Can W, Ming N, Yonghua S, Zhao X. Pareto Optimal Prediction Intervals of Electricity Price. IEEE transactions on power systems. 2017;32:817-9.
[16] Xiao C, Dong Z, Xu Y, Meng K, Zhou X, Zhang X. Rational and self-adaptive evolutionary extreme learning machine for electricity price forecast. Memetic computing. 2016;8:223-33.
[17] Shrivastava NA, Khosravi A, Panigrahi BK. Prediction Interval Estimation of Electricity Prices Using PSO-Tuned Support Vector Machines. IEEE transactions on industrial informatics. 2015;11:322-31.
[18] Nie Y, Liang N, Wang J. Ultra-short-term wind-speed bi-forecasting system via artificial intelligence and a double-forecasting scheme. Applied Energy. 2021;301:117452.
[19] Yousfi S, Berrani S-A, Garcia C. Contribution of recurrent connectionist language models in improving LSTM-based Arabic text recognition in videos. Pattern recognition. 2017;64:245-54.
[20] Cai M, Liu J. Maxout neurons for deep convolutional and LSTM neural networks in speech recognition. Speech communication. 2016;77:53-64.
[21] Zhang Y, Hutchinson P, Lieven NAJ, Nunez-Yanez J. Remaining Useful Life Estimation Using Long Short-Term Memory Neural Networks and Deep Fusion. IEEE access. 2020;8:19033-45.
[22] Sorin V, Barash Y, Konen E, Klang E. Deep Learning for Natural Language Processing in Radiology—Fundamentals and a Systematic Review. Journal of the American College of Radiology. 2020;17:639-48.
[23] Sundermeyer M, Ney H, Schlüter R. From feedforward to recurrent LSTM neural networks for language modeling. IEEE/ACM transactions on audio, speech, and language processing. 2015;23:517-29.

[24] Pang Z, Niu F, O'Neill Z. Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons. Renewable energy. 2020;156:279-89.
[25] Shi Z, Liang H, Dinavahi V. Direct Interval Forecast of Uncertain Wind Power Based on Recurrent Neural Networks. IEEE transactions on sustainable energy. 2018;9:1177-87.
[26] YU X-y, SHEN Y-x, CHEN J, JI Z-c. A Multi-Objective Prediction Method for Short-Term Microgrid Load Considering Interval Probability. ACTA ELECTONICA SINICA. 2017;45:930.
[27] Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997;9:1735-80.
[28] Zhang Z, Pan X, Jiang T, Sui B, Liu C, Sun W. Monthly and Quarterly Sea Surface Temperature Prediction Based on Gated Recurrent Unit Neural Network. Journal of marine science and engineering. 2020;8:249.
[29] de Simón-Martín M, Bracco S, Rosales-Asensio E, Piazza G, Delfino F, Giribone PG. Electricity Spot Prices Forecasting for MIBEL by using Deep Learning: a comparison between NAR, NARX and LSTM networks. 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe): IEEE; 2020. p. 1-6.
[30] Peng L, Liu S, Liu R, Wang L. Effective long short-term memory with differential evolution algorithm for electricity price prediction. Energy (Oxford). 2018;162:1301-14.
[31] Afrasiabi M, Mohammadi M, Rastegar M, Kargarian A. Probabilistic deep neural network price forecasting based on residential load and wind speed predictions. IET renewable power generation. 2019;13:1840-8.
[32] Lago J, De Ridder F, De Schutter B. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. Applied energy. 2018;221:386-405.
[33] Wei H, Hongxuan Z, Yu D, Yiting W, Ling D, Ming X. Short-term optimal operation of hydro-wind-solar hybrid system with improved generative adversarial networks. Applied Energy. 2019;250:389-403.
[34] Fan C, Sun Y, Zhao Y, Song M, Wang J. Deep learning-based feature engineering methods for improved building energy prediction. Applied energy. 2019;240:35-45.
[35] Takahashi S, Chen Y, Tanaka-Ishii K. Modeling financial time-series with generative adversarial networks. Physica A. 2019;527:121261.
[36] Chen Y, Wang X, Zhang B. An Unsupervised Deep Learning Approach for Scenario Forecasts. Power Systems Computation Conference; 2018. p. 1-7.
[37] Negnevitsky M, Mandal P, Srivastava AK. Machine Learning Applications for Load, Price and Wind Power Prediction in Power Systems. IEEE; 2009. p. 1-6.
[38] Yoon J. End-to-End Machine Learning Frameworks for Medicine: Data Imputation, Model Interpretation and Synthetic Data Generation. ProQuest Dissertations Publishing; 2020.
[39] Yoon J, Jarrett D, Van der Schaar M. Time-series generative adversarial networks. 2019.
[40] Uniejewski B, Marcjasz G, Weron R. On the importance of the long-term seasonal component in day-ahead electricity price forecasting. Energy economics. 2019;79:171-82.
[41] Jonsson T, Pinson P, Nielsen HA, Madsen H, Nielsen TS. Forecasting Electricity Spot Prices Accounting for Wind Power Predictions. IEEE transactions on sustainable energy. 2013;4:210-8.
[42] Yang Z, Ce L, Lian L. Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. Applied energy. 2017;190:291-305.
[43] Can W, Zhao X, Yelei W, Zhao Yang D, Kit Po W. A Hybrid Approach for Probabilistic Forecasting of Electricity Price. IEEE transactions on smart grid. 2014;5:463-70.
[44] Wang Y, Hug G, Liu Z, Zhang N. Modeling load forecast uncertainty using generative adversarial

networks. Electric power systems research. 2020;189:106732.
[45] Karami M, Shayanfar HA, Aghaei J, Ahmadi A. Scenario-based security-constrained hydrothermal coordination with volatile wind power generation. Renewable & sustainable energy reviews. 2013;28:726-37.
[46] Cui M, Ke D, Sun Y, Gan D, Zhang J, Hodge B-M. Wind Power Ramp Event Forecasting Using a Stochastic Scenario Generation Method. IEEE transactions on sustainable energy. 2015;6:422-33.
[47] Clements AE, Herrera R, Hurn AS. Modelling interregional links in electricity price spikes. Energy economics. 2015;51:383-93.
[48] Higgs H. Modelling price and volatility inter-relationships in the Australian wholesale spot electricity markets. Energy economics. 2009;31:748-56.
[49] Ignatieva K, Trück S. Modeling spot price dependence in Australian electricity markets with applications to risk management. Computers & operations research. 2016;66:415-33.
[50] Bell WP, Wild P, Foster J, Hewson M. Revitalising the wind power induced merit order effect to reduce wholesale and retail electricity prices in Australia. Energy economics. 2017;67:224-41.
[51] Forrest S, MacGill I. Assessing the impact of wind generation on wholesale prices and generator dispatch in the Australian National Electricity Market. Energy policy. 2013;59:120-32.
[52] Nazifi F, Nazifi F. The pass-through rates of carbon costs on to electricity prices within the Australian National Electricity Market. Environmental economics and policy studies. 2016;18:41-62.
[53] Higgs H, Worthington A. Stochastic price modeling of high volatility, mean-reverting, spike-prone commodities: The Australian wholesale spot electricity market. Energy economics. 2008;30:3172-85.
[54] Wang J, Du P, Hao Y, Ma X, Niu T, Yang W. An innovative hybrid model based on outlier detection and correction algorithm and heuristic intelligent optimization algorithm for daily air quality index forecasting. Journal of environmental management. 2020;255:109855.
[55] Bottou L, Arjovsky M, Lopez-Paz D, Oquab M. Geometrical Insights for Implicit Generative Modeling. Cham: Springer International Publishing; 2018. p. 229-68.
[56] Amjady N, Keynia F. A new prediction strategy for price spike forecasting of day-ahead electricity markets. Applied soft computing. 2011;11:4246-56.
[57] Nelson T, Kelley S, Orton F. A literature review of economic studies on carbon pricing and Australian wholesale electricity markets. Energy policy. 2012;49:217-24.
[58] Theil H. The analysis of disturbances in regression analysis. Journal of the American Statistical Association. 1965;60:1067-79.
[59] Moral-Carcedo J, Pérez-García J. Temperature effects on firms' electricity demand: An analysis of sectorial differences in Spain. Applied energy. 2015;142:407-25.
[60] Shao H, Deng X, Jiang Y. A novel deep learning approach for short-term wind power forecasting based on infinite feature selection and recurrent neural network. Journal of renewable and sustainable energy. 2018;10:43303.
[61] Peng C, Zhou H, Gu H, Yi X, Chi Y. A Power Market Nodal Price Forecasting Based on Historical Electricity Price and Unified Dispatch Load. 2021 IEEE 4th International Electrical and Energy Conference (CIEEC): IEEE; 2021. p. 1-8.
[62] Uniejewski B, Marcjasz G, Weron R. Understanding intraday electricity markets: Variable selection and very short-term price forecasting using LASSO. International Journal of Forecasting. 2019;35:1533-47.
[63] Diebold FX, Mariano RS. Com paring predictive accu racy. Journal of Business and Economic Statistics. 1995;13:253-63.

[64] Chen X, Dong ZY, Meng K, Xu Y, Wong KP, Ngan H. Electricity price forecasting with extreme learning machine and bootstrapping. IEEE Transactions on Power Systems. 2012;27:2055-62.
[65] Shen Y, Lu X, Yu X, Zhao Z, Wu D. Short-term wind power intervals prediction based on generalized morphological filter and artificial bee colony neural network. 2016 35th Chinese Control Conference (CCC): IEEE; 2016. p. 8501-6.
[66] Aprillia H, Yang H-T, Huang C-M. Statistical load forecasting using optimal quantile regression random forest and risk assessment index. IEEE Transactions on Smart Grid. 2020;12:1467-80.
[67] Lotfi M, Javadi M, Osório GJ, Monteiro C, Catalão JP. A novel ensemble algorithm for solar power forecasting based on kernel density estimation. Energies. 2020;13:216.

Appendix A. Autocorrelation analysis of residuals

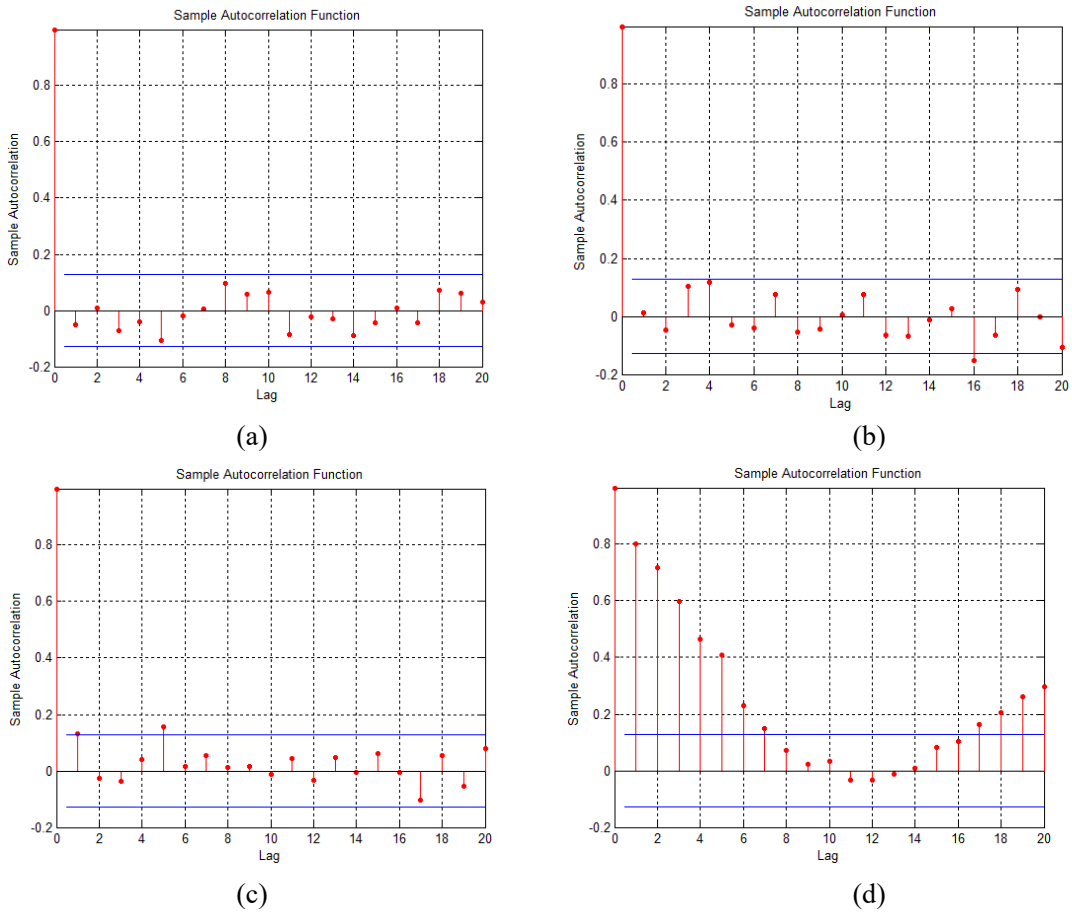


Fig. A1 ACF test of residuals with different models
(a) CTSGAN, (b) LSTM, (c) LASSO and (d) BPNN