

Article

A Prosumer Power Prediction Method Based on Dynamic Segmented Curve Matching and Trend Feature Perception

Biyun Chen ^{1,*}, Qi Xu ¹, Zhuoli Zhao ², Xiaoxuan Guo ³, Yongjun Zhang ⁴, Jingmin Chi ⁵ and Canbing Li ⁶

¹ Key Laboratory of Power System Optimization and Energy Saving Technology, Guangxi University, Nanning 530004, China

² Department of Electrical Engineering, School of Automation, Guangdong University of Technology, Guangzhou 510006, China

³ Electric Power Research Institute, Guangxi Power Grid Corporation, Nanning 530023, China

⁴ School of Electric Power, South China University of Technology, Guangzhou 510640, China

⁵ Guangxi Minhai Energy Co., Ltd., Nanning 530012, China

⁶ Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

* Correspondence: chenbiyu@gxu.edu.cn

Abstract: With the massive installation of distributed renewable energy (DRE) generation, many prosumers with the dual attributes of load and power supply have emerged. Different DRE permeability and the corresponding peak-valley timing characteristics have an impact on the power features of prosumers, so new models and methods are needed to reflect the new features brought about by these factors. This paper proposes a method for predicting the power of prosumers. In this method, dynamic segmented curve matching is applied to reduce the complexity of source-load coupling features and improve the effectiveness of the input features, and trend feature perception based on a temporal convolutional network (TCN) was applied to grasp the power trend of prosumers by predicting the multisegment trend indexes. The LST-Atten prediction model based on a temporal attention mechanism (TAM) and a long short-term memory (LSTM) network was applied to predict “day-ahead” power, which combines the trend indexes and similar curve sets as the input. Simulation results show that the proposed model has higher accuracy than individual models. Furthermore, the proposed model can maintain prediction stability under different renewable energy permeability scenarios.

Keywords: dynamic segmented curve matching; LST-Atten; power prosumer; power prediction; trend feature perception



Citation: Chen, B.; Xu, Q.; Zhao, Z.; Guo, X.; Zhang, Y.; Chi, J.; Li, C. A Prosumer Power Prediction Method Based on Dynamic Segmented Curve Matching and Trend Feature Perception. *Sustainability* **2023**, *15*, 3376. <https://doi.org/10.3390/su15043376>

Academic Editors: Chun Wei, Miltiadis (Miltos) Alamaniotis, Xiaoqing Bai, Peijie Li and Dongliang Xiao

Received: 1 December 2022

Revised: 26 January 2023

Accepted: 10 February 2023

Published: 12 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the improvement of DRE generation technology, many power consumers turn into prosumers [1,2]. However, prosumer power will have strong uncertainty on both the source and load sides because of the integration of DRE and the implementation of demand-response policies. At the same time, the variation in power will be complicated by prosumers in terms of the bidirectional power flow and fluctuation of DRE generation [3]. Therefore, there is an urgent need to explore new power prediction methods to deal with the double uncertainty on both the source and load sides.

At present, there are two kinds of power prediction methods for DRE [4]: decoupling and direct prediction. The decoupling prediction method first decouples prosumer power into wind power, photovoltaic (PV) power, and load. Then, different prediction models are used to learn the different features of the DRE generation and load. Finally, the above models' results are integrated to obtain the predicted power. In this regard, reference [5] proposes a feature extraction method based on a prosumer power curve to optimize pairing, which estimates the capacity of the distributed photovoltaic systems (DPVS) by an integrated model based on multiple support vector regression. Reference [6]

proposes a linear-based estimator to separate the power generation from the feeder-level measurements using the measured values of the substation and the measured power of a nearby PV plant. However, the drawbacks of the above methods are heavy reliance on PV generation data and a lack of DPVS information. Furthermore, after the decoupling of prosumer power, the accuracy needs to be improved via the accurate prediction of each component.

The direct prediction method can be broadly divided into traditional methods and intelligence-based methods. Traditional methods include multiple linear regression [7], exponential smoothing [8], and autoregressive moving averages [9]. Intelligence-based methods include support vector machines (SVM) [10], artificial neural networks [11], deep neural networks [12], and hybrid algorithms. Reference [13] proposes an integrated genetic algorithm (GA) and bidirectional gated recurrent unit (Bi-GRU) hybrid data-driven technique for short-term-load forecasting. Reference [14] proposes a fusion model based on a light gradient boosting machine (Lightgbm) and LSTM to forecast short-term photovoltaic power generation. A prediction model based on the integration of deep neural network and wavelet transformation is used to improve the net power prediction accuracy in reference [15]. Bayesian deep learning was used in reference [16] to capture stochastic uncertainty, and this achieved better results in power prediction. Reference [17] combines the TAM with time-series characteristics and proposes a prediction model based on TCN. Direct prediction methods have higher prediction accuracy than decoupling prediction methods when the model is properly chosen [18]. Therefore, the challenge is how to integrate all the relevant factors into the model in a reasonable way while accurately reflecting the cyclical load features and the stochastic features of DRE generation.

On the one hand, as DRE generation and load are closely related to meteorological factors, such as wind speed, temperature, solar radiance, and humidity, there is a certain coupling relationship between the sources and loads. Due to the insufficient consideration of the dynamic coupling relationship between the sources and loads in the existing studies, it is difficult to cope with the impact on the power system brought by the gradually increasing uncertainty of the sources and loads. If this coupling relationship can be taken into account, the prediction accuracy can be effectively improved. On the other hand, the increasingly fluctuating power prediction needs to pay more attention to the local variation from the prosumers' side, but the single-structured model lacks the ability to learn the higher-order features such as the time-sharing features and trend features, which determines whether the prediction accuracy can be improved. Accordingly, it is important to construct a corresponding feature analysis model to improve the ability of feature extraction and prediction.

For the above difficulties, we propose a power prediction method that aims to learn the time-sharing features and trend features of the power curves. After focusing on the coupling relationship between the source and load in different DRE penetration scenarios, the proposed method achieves accurate predictions directly, which can effectively avoid the accumulation of prediction errors generated by the decoupling power.

The key contributions can be summarized as follows:

- We establish a short-term power prediction model based on dynamic curve segmentation and trend feature perception, which combines SVM, TCN, and LST-Atten algorithms;
- We simulate and evaluate the prediction performance of the proposed model in three different permeability scenarios. The prediction accuracy of the proposed model can maintain high prediction accuracy under different permeability scenarios;
- We design three comparative simulations to verify the effectiveness of dynamic curve segmentation and trend feature perception. The simulation results show that the proposed model has a better prediction effect than other models.

The remainder of the paper is organized as follows. Section 2 describes the definition of prosumer power and analyzes the source-load coupling features. The methods and steps of power-curve clustering and dynamic segmentation are introduced in Section 3. Section 4 introduces the methods and details of the trend feature perception module.

Section 5 proposes the framework and details of the prediction model. Section 6 explains the simulations and corresponding results, while Section 7 concludes the paper.

2. Analysis of Power Features

The power of prosumers has different characteristics from that of pure consumers. It is not a simple linear superposition of the power of the DRE and the load but exhibits a complex dynamic coupling feature, which needs a new modeling structure to accurately reflect it.

2.1. Definition of Prosumer Power

The prosumer power can be defined as the actual load minus the DRE generation, as shown in the following equation [19]:

$$P_n = P_u - P_s \quad (1)$$

where P_n is the prosumer power, P_u is the actual load, and P_s is the DRE generation.

2.2. Analysis of the Coupling Features

Prosumer power will be affected not only by load changes but also by DRE generation. Specifically, the load changes are related to factors such as consumption behavior, weather, and day types, while DRE generation will be affected by equipment parameters and meteorological factors such as solar radiance and wind speed. Therefore, the factors affecting the load and the DRE generation will also affect the prosumer power. In this case, the coupling features of DRE generation and load should be taken into account when forecasting the power of prosumers. To visualize the coupling features between DRE generation and load, we increase the DRE penetration in the Tempe campus of Arizona State University [20] to 50% to simulate a high penetration scenario; the specific curves are plotted for typical days in summer and winter as shown in Figure 1.

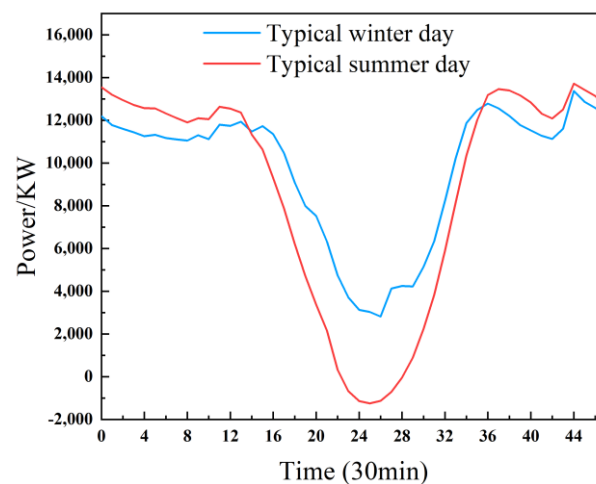


Figure 1. Power curves in summer and winter.

As seen in Figure 1, the power curves generated by the coupling features of DRE generation and load show an obvious “duck curve”. The high DRE penetration increases the daily peak-to-valley difference, and the PV generation at noon will reduce the power significantly. But in the evening, the load demand rises, and the PV generation weakens, resulting in a sudden rise in power demand. By comparing the curve patterns of typical days in summer and winter, it can be seen that the power curves show different trend patterns at different periods. The peak-to-valley difference in summer is significantly larger than that in winter, and the “concave” period in summer is longer than that in winter. These indicate that the power curves have obvious seasonal and time-sharing features,

which are the results of the coupling of the DRE generation timeliness and the time-sharing consumption features. It is not difficult to conclude that the dominant factor affecting the power will not be a single fixed meteorological feature or historical power value under the effect of coupling features in different periods, which puts high requirements on the feature learning ability of the prediction models.

3. Feature Matching of the Power Curves

As mentioned above, the power of prosumers has the feature of both seasonality and timeliness caused by DRE generation and the time-sharing consumption features caused by consumers. Therefore, feature matching and trend index prediction are proposed to reduce the complexity of source-load coupling, which are discussed in Sections 3 and 4, respectively.

3.1. Curve Clustering Considering Power Feature Indexes

Power features are visualized by power curve types. However, if the clustering algorithm based on Euclidean distance is used, the power curves with different features may be classified into the same type, which affects the quality of the clustering and reduces the training effect of the prediction model. Therefore, we introduce five daily power feature indexes to cluster the power curves, which include daily load factor, daily peak-to-valley difference, maximum power utilization time, daytime (7:30–19:30, total 12 h) load factor, and nighttime (0:00–7:30, 19:30–24:00, total 12 h) load factor. Meanwhile, the entropy weight method is used to calculate the weight coefficients of daily power feature indexes, which reflect the importance of each index to characterize the power curve.

The specific steps are as follows:

Step 1: Daily normalization of the power curve to eliminate the effect of natural growth.

Step 2: Calculate the entropy value of the daily power feature indexes, as shown in the following equations [21]:

$$h_j = -e \sum_{i=1}^n f_{ij} \ln f_{ij} \quad (2)$$

$$f_{ij} = \frac{r_{ij}}{\sum_{i=1}^n r_{ij}} \quad (3)$$

$$e = 1/\ln n \quad (4)$$

where h_j is the entropy value, e is the standardized coefficient, $i \in \{1, 2, \dots, n\}$, n is the number of power curves to be clustered, $j \in \{1, 2, \dots, m\}$, m is the number of daily power feature indexes, r_{ij} is daily power feature indexes data, and f_{ij} is the degree of contribution.

Step 3: Calculate the entropy weight ω_j of each daily power feature index [21]:

$$\omega_j = \frac{\exp\left(\sum_{t=1}^n h_t + 1 - h_j\right) - \exp(h_j)}{\sum_{l=1}^m \left(\exp\left(\sum_{t=1}^n h_t + 1 - h_l\right) - \exp(h_l)\right)} \quad (5)$$

where $t \in \{1, 2, \dots, m\}$, $l \in \{1, 2, \dots, m\}$.

Step 4: Calculate the Euclidean distance between each index and the cluster center, and then multiply the entropy weight to get the improved Euclidean distance.

Step 5: The clustering method used the K-means algorithm based on improved Euclidean distance and evaluated by silhouette coefficient [22]. The silhouette coefficient of single data point i shows in (6), which indicates how tightly grouped the data points are in

that cluster. The larger the silhouette coefficient, the better the clustering effect. Naturally, the k with the largest average silhouette coefficient is the optimal number of clusters.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (6)$$

where $b(i)$ is the average distance between point i and all samples in the nearest cluster and $a(i)$ is the average distance between point i and other samples in the same cluster.

The specific clustering process of K-means is shown in Figure 2:

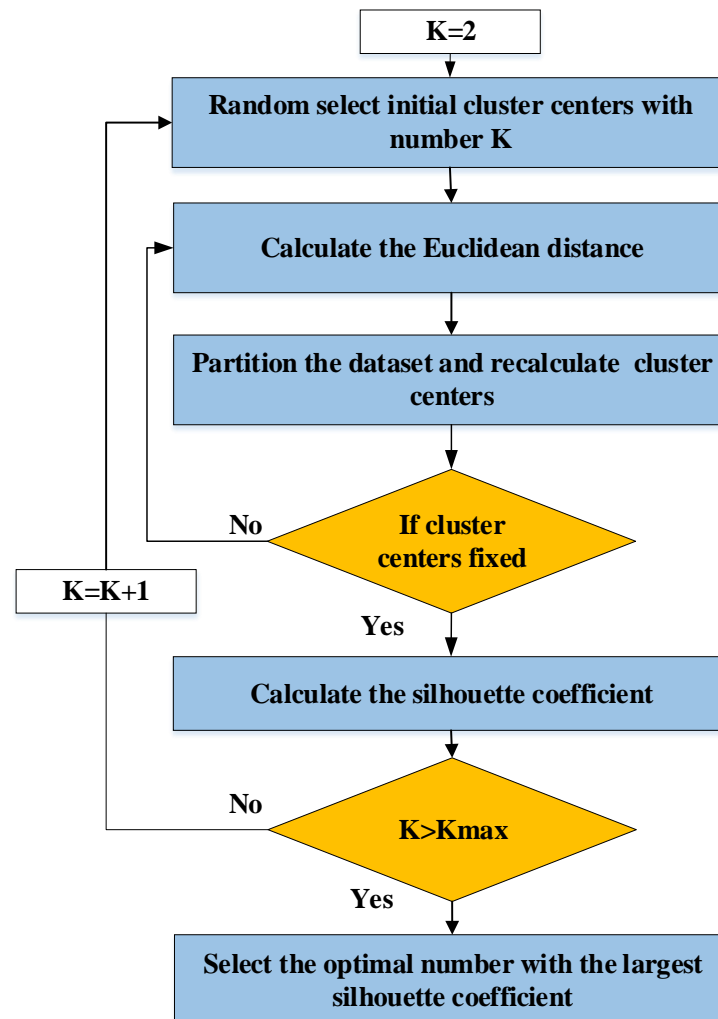


Figure 2. Flowchart of K-means.

3.2. Dynamic Segmentation of Power Curves

The power curve fluctuations have a certain daily periodicity, which shows that the daily power peaks and valleys are located roughly in the same period. According to this rule and the center curves of clusters, the power curve can be dynamically segmented. Different segments have different power curve fluctuations and dominant influencing factors. Thus, the dominant factors affecting the power variation need to be selected for different periods.

When taking a PV-oriented campus as an example, the steps of dynamic segmentation are as follows:

Step 1: Select sunrise and sunset times as the sunshine segmentation points based on solar radiation data.

Step 2: Calculate the power change rate for each cluster center curve:

$$\lambda_1(i) = \frac{x_i - x_{i-1}}{x_i} \times 100\%, \lambda_2(i) = \frac{x_{i+1} - x_i}{x_i} \times 100\% \quad (7)$$

where $\lambda_1(i)$ and $\lambda_2(i)$ are the adjacent power change rates for point i , x_i is the power for point i .

Step 3: Judge whether it is an inflection point: If all central curves at point i satisfy $\lambda_1(i)\lambda_2(i) < 0$, point i is an inflection point, then proceed to step 4.

Step 4: Calculating the relative change rate $\lambda'(i)$ of the inflection point:

$$\lambda'(i) = |\lambda_1(i) - \lambda_2(i)| \quad (8)$$

Step 5: Select the inflection point with the largest relative change rate as the trend mutation segmentation point.

The dynamic segmentation results are shown in Figure 3. In Figure 3, points 14 and 34 are sunshine segmentation points. Points 4, 5, 36, 42, and 44 are inflection points, and the corresponding average relative change rates are 6.78, 5.81, 4.01, 22.56, and 13.62%, respectively. According to the calculation results, we choose inflection point 42 as the trend mutation segmentation point.

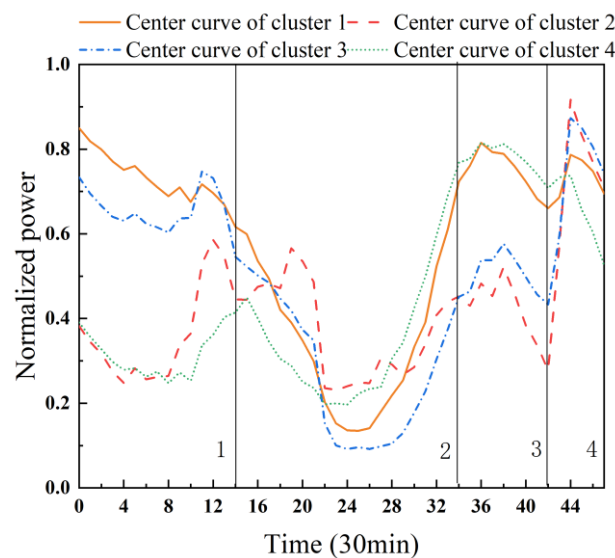


Figure 3. Dynamic segmentation results.

4. Trend Indexes Prediction Module

In most prediction models using trend indexes, fixed trend indexes, such as historical power growth rates over the same period, are selected. In this paper, we divide the multiple segments that can reflect the power features and then select a few representative power values as trend indexes. A temporal convolutional network (TCN) algorithm was used to predict the trend indexes.

4.1. Selection of Trend Indexes and Feature Dimension Screening

Based on the results of the power curve segmentation, assuming the number of segments is four, the maximum power P_{\max} , minimum power P_{\min} , average power P_{av} , gross power P_{sum} , and average power of the four segments ($P_{av.1}$, $P_{av.2}$, $P_{av.3}$, $P_{av.4}$) are selected as trend indexes for prediction.

Before predicting the trend indexes, numerous meteorological features need to be screened for dominant factors and temporal dimensions, and irrelevant or low-correlation meteorological features and temporal dimensions need to be eliminated. We use Pearson

correlation coefficients (PCCs) to characterize the correlation between trend indexes and influencing factors. The PCCs are calculated as

$$r_{xy} = \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^m (y_i - \bar{y})^2}} \quad (9)$$

where \bar{x} and \bar{y} are the average values of the elements in each vector.

4.2. Structure of TCN

TCN is a convolutional neural network (CNN) architecture optimally adapted to solve time-series problems, which introduces dilated causal convolution and residual blocks on top of CNN. The structure of TCN is shown in Figure 4.

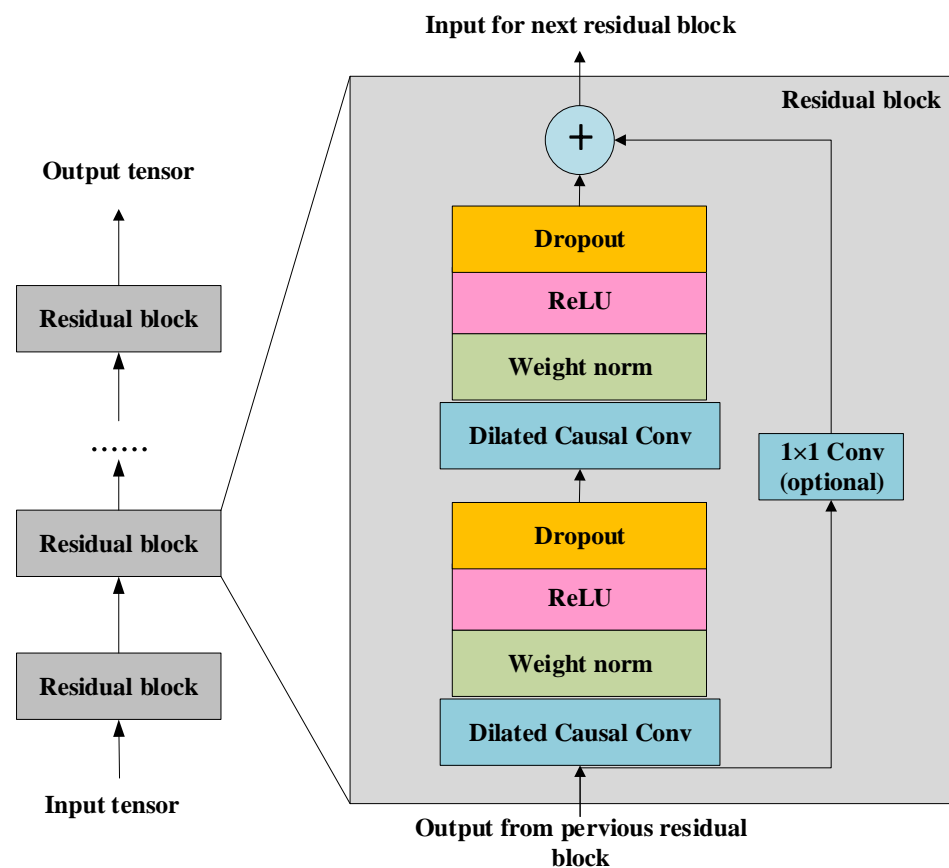


Figure 4. Structure of TCN.

4.3. Trend Indexes Prediction Process

When building the trend indexes prediction module, meteorological data of different time scales are used as inputs for the trend indexes of different periods. The specific implementation steps are as follows:

Step 1: Calculate the PCCs between each trend index and the influencing factors, and then screen for dominant factors affecting each trend index.

Step 2: Calculate the PCCs between segmented trend indexes and the time dimensions of dominant meteorological factors, and then exclude low correlation time dimensions.

Step 3: Establish the corresponding TCN model for each trend index separately, input the filtered influencing factors, and finally output the trend indexes prediction results.

The flow chart of the trend indexes prediction module is shown in Figure 5.

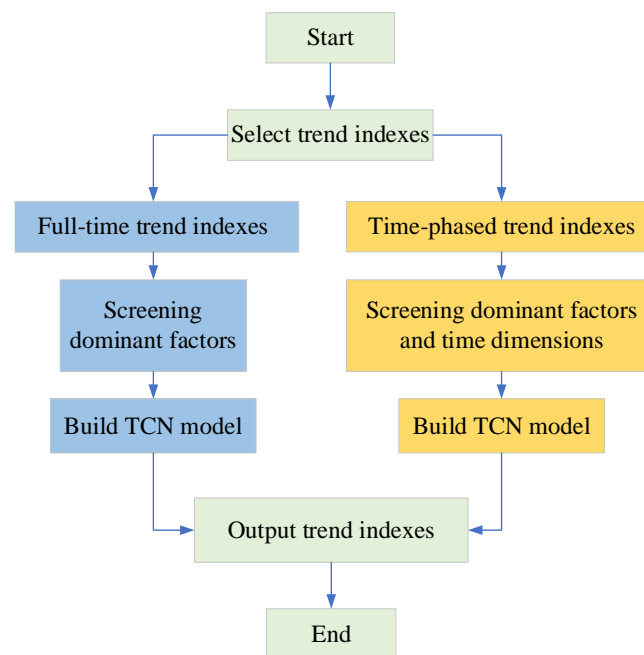


Figure 5. Flow chart of trend indexes prediction module.

5. Proposed Model

5.1. Model Design

The proposed prediction model includes three modules: the trend indexes prediction module, the curve similarity matching module, and the short-term power prediction module.

The main steps of the proposed model are as follows:

Step 1: Data preprocessing. To eliminate the difference in magnitude and avoid the gradient problem during model training, the following equation is used to normalize the data of various features, including power and weather [23–25].

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (10)$$

where x is the data to be normalized, x^* is the normalized data, x_{\max} is the maximum value, and x_{\min} is the minimum value.

Step 2: Power curve clustering and dynamic segmentation. Clustering curves based on historical power data, and then curve dynamic segmentation based on the clustering results.

Step 3: Trend indexes prediction module. Extract and forecast the multiperiod trend indexes.

Step 4: Curve similarity matching module. When combining the results of the dynamic segmentation curve and trend index prediction, the similar set of the power curve is selected.

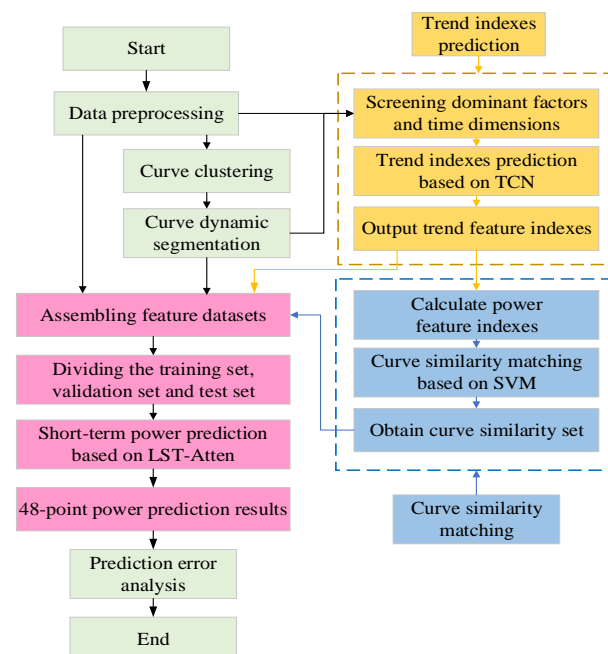
Step 5: Build a short-term power prediction model based on LST-Atten. We select trend indexes, similar curve sets, and meteorological data as inputs. The inputs used to forecast the 48-point power of the next day are listed in Table 1.

Table 1. Inputs for the power prediction model.

Size	Input	Note
48	GHI	Global horizontal irradiance (w/m^2)
48	CDNI	Clear sky direct normal irradiance
48	CDHI	Clear sky diffuse horizontal irradiance
48	SZA	Solar zenith angle (Degree)
48	Wind speed	m/s
48	Dew point	$^{\circ}C$
48	Water	Precipitable water (/cm)
48	Temperature	$^{\circ}C$
144	Historical power of the first three days	Historical data
1	Day type	Weekday: 0; Weekend: 1
1	Season	One-hot code
1	Maximum power	From trend indexes prediction
1	Minimum power	From trend indexes prediction
1	Average power	From trend indexes prediction
1	Gross power	From trend indexes prediction
4	Average power of four periods	From trend indexes prediction

Step 6: Divide the dataset into the first 75% as training data, 15% as validation data, and the last 10% as test data. With the goal of minimizing the loss function in the validation data, hyper-parameters have been set to achieve optimal solutions. Finally, the test data is used to evaluate the performance of the final model.

The flow chart is shown in Figure 6.

**Figure 6.** Flow chart of the power prediction model.

5.2. Curve Similarity Matching Module

The traditional method selects similar curves by calculating the similarity of the evaluation vector, which includes day type and meteorological data. In this paper, we propose a dynamic segmentation matching method considering power feature indexes and meteorological data. The power characteristic index is calculated from the trend indexes of the trend indexes prediction module. The SVM classification algorithm is used to replace

the process of similarity calculation and directly predict the class of similar curves. The power feature indexes are defined as shown in Table 2.

Table 2. Power feature indexes.

Time Interval	Index	Definition
00:00–24:00	Load factor	$a_1 = P_{av} / P_{max}$
00:00–24:00	Maximum power utilization time	$a_2 = P_{sum} / P_{max}$
00:00–24:00	Peak-to-valley difference ratio	$a_3 = \frac{(P_{max} - P_{min})}{P_{max}}$
0:00~7:00	Segment 1 load factor	$a_4 = P_{av.1} / P_{av}$
7:00~17:00	Segment 2 load factor	$a_5 = P_{av.2} / P_{av}$
17:00~21:00	Segment 3 load factor	$a_6 = P_{av.3} / P_{av}$
21:00~24:00	Segment 4 load factor	$a_7 = P_{av.4} / P_{av}$

The specific steps of the curve similarity matching module are as follows:

Step 1: Obtain the trend indexes from the trend indexes prediction module, and calculate the power feature indexes according to the definition.

Step 2: Select the historical meteorological data and power feature indexes as training inputs. The expected output is the corresponding similar curve categories. Then put the data into the SVM model for training, set the number of iterations to 300, and save the trained model.

Step 3: Input the real-time meteorological data and power feature indexes of the day to be predicted into the trained model, and obtain similar sets of curves.

Step 4: Evaluate the effect of similar curve sets. We introduce accuracy and morphological similarity distance [26] D_{MSD} as the evaluation metrics of similar curve selection. D_{MSD} can measure both the numerical spacing of the curves and the similarity of the curve shapes. The more similar the curve shape, the smaller the value of D_{MSD} . The relevant equations are as follows:

$$D_{MSD}(L_c, L_f) = \sqrt{\sum_{k=1}^n (l_{c,k} - l_{f,k})^2} \left(2 - \frac{\left| \sum_{k=1}^n (l_{c,k} - l_{f,k}) \right|}{\sum_{k=1}^n |l_{c,k} - l_{f,k}|} \right) \quad (11)$$

$$L_c = [l_{c,1}, l_{c,2}, \dots, l_{c,k}, \dots, l_{c,n}], L_f = [l_{f,1}, l_{f,2}, \dots, l_{f,k}, \dots, l_{f,n}] \quad (12)$$

$$A_{cc} = \frac{T}{S} \quad (13)$$

where L_c are the power curves series to be compared, L_f are the actual power curves series, n is the length of the power curves series, T is the number of correctly classified samples, and S is the number of samples.

5.3. LST-Atten

LST-Atten consists of TAM and LSTM networks [27]. TAM is a mechanism that mimics the allocation of attentional resources in the human brain by focusing on temporally important features from a large amount of information. The introduction of TAM aims to enhance the LSTM model's memory of long-time series information, highlight the key temporal factors, and improve the model prediction effect.

6. Case Study

The public datasets of Tempe, Downtown, and Polytechnic campuses on the website of the Campus Metabolism program [20] are used to verify the performance and feasibility of the proposed model. The datasets contain the electric load and DRE generation data from 1 January 2018 to 31 December 2019. The environmental data were chosen from the weather station closest to each campus and were downloaded from the National Solar

Radiation Data Bank website [28]. The data include temperature, dew point, humidity, wind speed, global horizontal irradiance (GHI), clear sky direct normal irradiance (CDNI), clear sky diffuse horizontal irradiance (CDHI), and solar zenith angle (SZA), etc. The data resolution is half an hour.

6.1. Analysis of Influencing Factors

Existing studies consider the impact of different DRE penetration on the power variation less. The three campuses selected in this paper have different DRE penetrations, which are set as different scenarios. Downtown in scenario 1 has a DRE penetration of 1.1%, Tempe in scenario 2 has a DRE penetration of 13%, and Polytechnic in scenario 3 has a DRE penetration of 21.7%. We selected meteorological data, such as GHI, CDNI, CDHI, SZA, etc., as the influencing factors. The PCCs between each influencing factor and power are calculated separately for three different DRE penetration scenarios. The calculated results are shown in Figure 7.

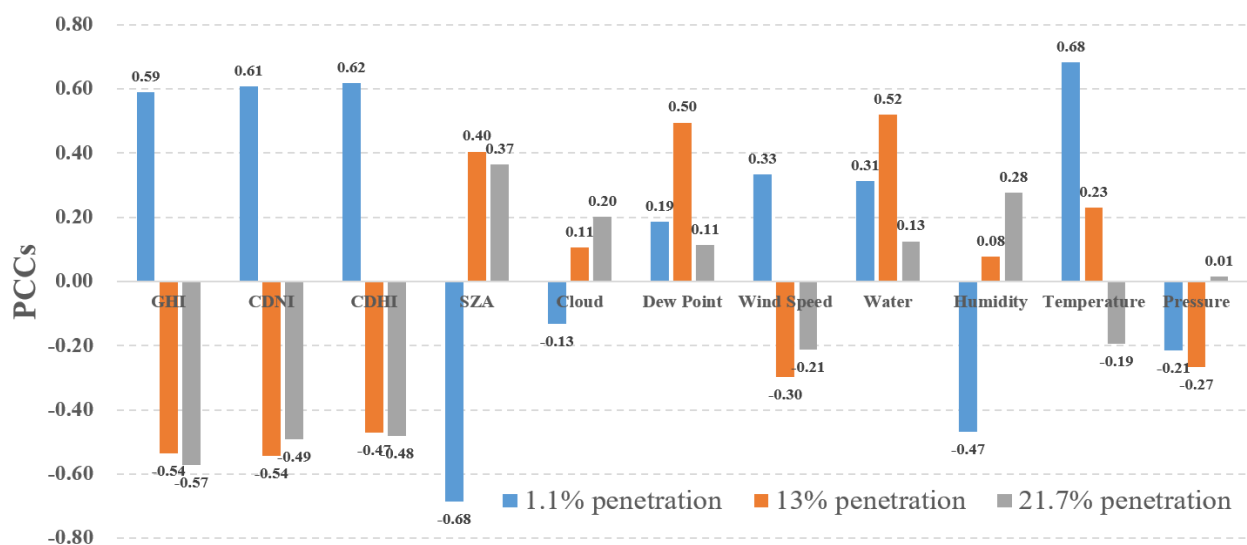


Figure 7. Analysis results of the influencing factors under different DRE penetration scenarios.

6.2. Measuring Metrics

The normalized root mean square error (NRMSE) [29], mean bias error (MBE), and mean absolute percentage error (MAPE) [30] were selected to measure the accuracy of the prediction model.

$$e_{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{x(i) - y(i)}{x(i)} \right| \times 100\% \quad (14)$$

$$e_{NRMSE} = \frac{1}{x_{\max}} \sqrt{\frac{1}{n} \sum_{i=1}^n (x(i) - y(i))^2} \quad (15)$$

$$e_{MBE} = \frac{1}{n} \sum_{i=1}^n x(i) - y(i) \quad (16)$$

where $x(i)$ is the actual value, $y(i)$ is the predicted value, x_{\max} is the actual maximum value, and n is the number of samples in the test set. They directly relate to accuracy: the smaller the NRMSE and MAPE are, the higher the accuracy. It is noteworthy that a positive MBE means that the model underestimates the power value, while a negative MBE means an overestimation.

6.3. Simulation Design

We use the following three simulations to verify the effectiveness and generalizability of the proposed power prediction method.

6.3.1. Comparative Simulation under Different Similar Sets Selection Method

In this section, scenario 2, with 13% penetration, was used as the simulation object. We use the traditional meteorological similar method and the dynamic segment matching method to select similar sets from December 1 to 30, 2019. A comparison of the curve similarity evaluation metrics is shown in Table 3.

Table 3. Comparison of curve similarity evaluation metrics.

Similar Sets Selection Method	Acc (%)	D _{MSD}
Meteorological Similar Method	— ¹	0.8987
Dynamic Segment Matching Method	80	0.7526

¹ Clustering curve types have no reference to the meteorological similarity method selection results.

Power prediction is performed based on the different similar sets of the power curves. We took the overall average of the measuring metrics for one month as the prediction results. The results of the measuring metrics are listed in Table 4.

Table 4. Comparison of measuring metrics under different similar sets selection methods.

Similar Sets Selection Method	MAPE (%)	NRMSE
Meteorological Similar Method	7.44	0.04792
Dynamic Segment Matching Method	6.09	0.04372

For a more detailed display, we choose 13 to 15 December 2019, for the typical day analysis, where the meteorological conditions on 13 and 14 are similar, but the diurnal temperature difference on 15 is large. Moreover, 13 is a weekday, whereas 14 and 15 are weekends. The prediction for these days should consider not only the meteorological changes but also the impact of the time-sharing consumption features under different day types. The prediction results are shown in Figure 8.

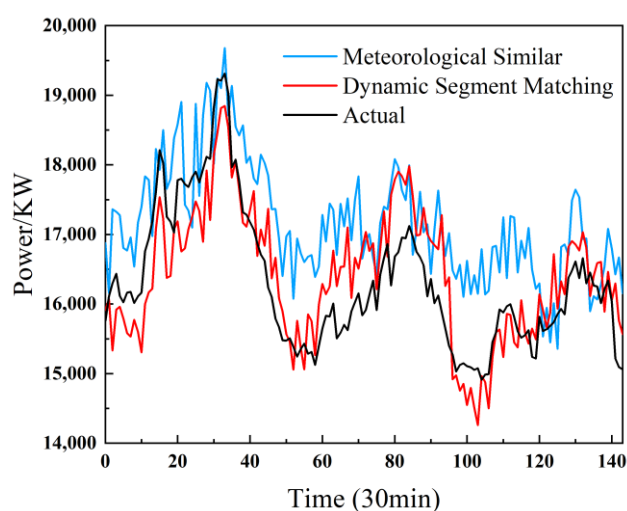


Figure 8. Prediction results under different similar set selection methods.

Obviously, a similar set with a high similarity of meteorological features does not mean its power curve shape is also similar because the power of prosumers contains random fluctuating components and the diversity of the influencing factors. The prediction performance based on the meteorological similarities method is not effective, especially for the inflection points. The comparison results show that the proposed method can learn the time-sharing consumption features of the prosumers through the dynamic segmentation.

6.3.2. Ablation Study

In order to verify the effectiveness of the dynamic segmentation module and the trend feature component on the power prediction model, we designed two models involved in the ablation study:

- Complete model: inputs contain complete trend indexes;
- Comparison model: inputs without time-phased trend indexes and the selection of the similarity set without considering segmented power feature indexes.

Scenario 3, with a 21% penetration, was used as the simulation object. We choose five consecutive days from November 11 to 15, 2019, as the simulation period, where 11 to 14 are weekdays with mostly sunny and cloudy weather, whereas 15 is the weekend with cloudy and rainy weather. The ablation study results are shown in Figure 9, focusing on the comparison of the prediction effects at key points for sudden changes in trends.

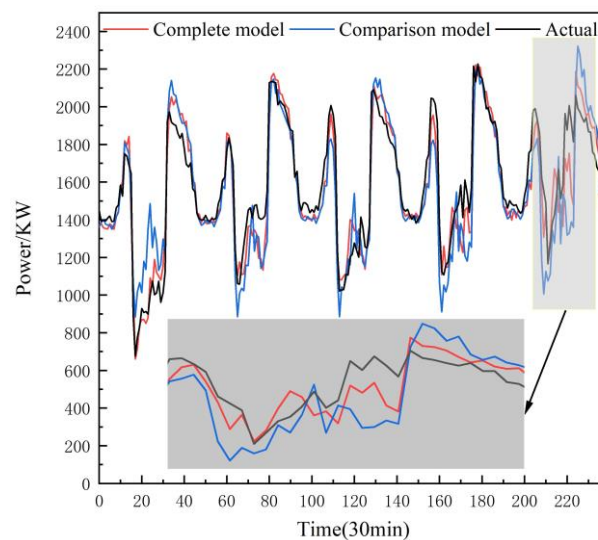


Figure 9. Prediction results of the ablation study.

As is visible from Figure 9, the comparison model without the time-phased prediction component is not effective in predicting the DRE generation fluctuation points. Furthermore, with the enlarged details in Figure 9, it is easy to see that the complete model with the time-phased prediction component has a significant improvement in the prediction of the occurrence time of the abrupt trend change points and the segmented peak and valley values. It is not difficult to conclude that the dynamic segmentation and the corresponding trend feature components help the prediction model learn the time-sharing consumption features and trend features by matching the source–load coupling features, which effectively improves the accuracy of the prediction.

6.3.3. Comparative Simulation with Different Power Prediction Models

In order to verify the superiority of the proposed model, we selected TCN, LSTM, and Lightgbm for a comparative simulation with the proposed model. More specifically, the TCN model consists of three layers of residual units and fully connected layers. The LSTM model consists of two hidden layers with 24 neurons. The iteration times of the above deep learning models are set to 200, the activation function of TCN is ReLU, and the activation function of LSTM is Sigmoid. The objective of the Lightgbm model is set to regression; the training method adopts a gradient lifting decision tree (GBDT); the number of iterations is 2000, and the other adjustable parameters are determined by the Adam algorithm.

Six consecutive months from July to December 2019 are selected as the simulation period. The results of the measuring metrics are listed in Table 5.

Table 5. Comparison of the measuring metrics for different DRE penetration scenarios.

Model	Scenario 1 (1.1% Penetration)			Scenario 2 (13% Penetration)			Scenario 3 (21.7% Penetration)		
	MAPE (%)	MBE (KW)	NRMSE	MAPE (%)	MBE (KW)	NRMSE	MAPE (%)	MBE (KW)	NRMSE
Proposed model	5.52	−25.43	0.04128	5.23	−382.63	0.04018	9.62	10.14	0.02806
TCN	6.45	−47.23	0.05213	6.84	−639.62	0.05102	10.34	−12.81	0.04431
LSTM	7.13	−37.92	0.05325	7.17	−811.58	0.05195	10.73	20.23	0.04539
Lightgbm	6.72	−42.41	0.05233	6.53	−706.87	0.05046	11.37	−24.56	0.06019

Table 5 illustrates the proposed model; TCN and LSTM have better prediction stability for different DRE penetration scenarios, while Lightgbm shows poor prediction adaptability for those scenarios with a high DRE penetration. When compared with TCN, which performs better among the three conventional models, the proposed model achieves the best performance, with further improvements of 14.41% and 20.81% for the MAPE and NRMSE, respectively, in scenario 1 with low DRE penetration. Similarly, the proposed model provides 6.96% and 36.68% less MAPE and NRMSE, respectively, compared to TCN in scenario 3 with higher DRE penetration. As can be seen from the MBE, the models tend to overestimate the power values in scenario 1 and scenario 2, while the proposed model will underestimate the power values in scenario 3. Additionally, note that the MAPE will be very large and make no sense in high penetration scenarios because the power may be very small when a highly DRE output is very high and the actual load is low.

July and December for scenario 2 are selected as typical months for more detail. The results of the measuring metrics are listed in Table 6.

Table 6. Comparison of measuring metrics for typical months in 13% penetration Scenario.

Model	July			December		
	MAPE (%)	MBE (KW)	NRMSE	MAPE (%)	MBE (KW)	NRMSE
Proposed model	5.08	174.52	0.03983	5.22	−536.83	0.04004
TCN	6.46	−625.21	0.05078	6.73	−861.93	0.05115
LSTM	7.05	−671.57	0.05299	7.11	−880.68	0.05306
Lightgbm	6.26	−825.56	0.04976	6.22	−906.87	0.04990

As can be seen from Table 6, the prediction effects of TCN, LSTM, and Lightgbm are not outstanding, although they are stable in the case of seasonal change. The MAPE and NRMSE metrics demonstrate that the proposed model has a significant enhancement effect on the power prediction effect of medium and high DRE penetration.

Furthermore, to verify the enhancement effect of the proposed model on local trend prediction, we selected the typical summer days of 22 to 24 July 2019, in scenario 1 and scenario 3 for simulation analysis. The prediction results are shown in Figures 9 and 10.

From Figure 10a, it can be seen that, in scenario 1, with low DRE penetration, the curve of the power shows a regular double-peak pattern. Although the PV generation during the midday leads to a certain magnitude of power fluctuation, the peak-to-valley difference and the time of peak-to-valley occurrence are mainly determined by consumption features. The dashed lines in Figure 10b are the dynamic segmented lines, which show the important points in the daily power curve of the sudden rise and sudden fall. Figure 10a,b demonstrate the proposed model is better than the three comparison models in predicting the peak and valley values and the curve trend change points, especially for midday when the power fluctuations are large.

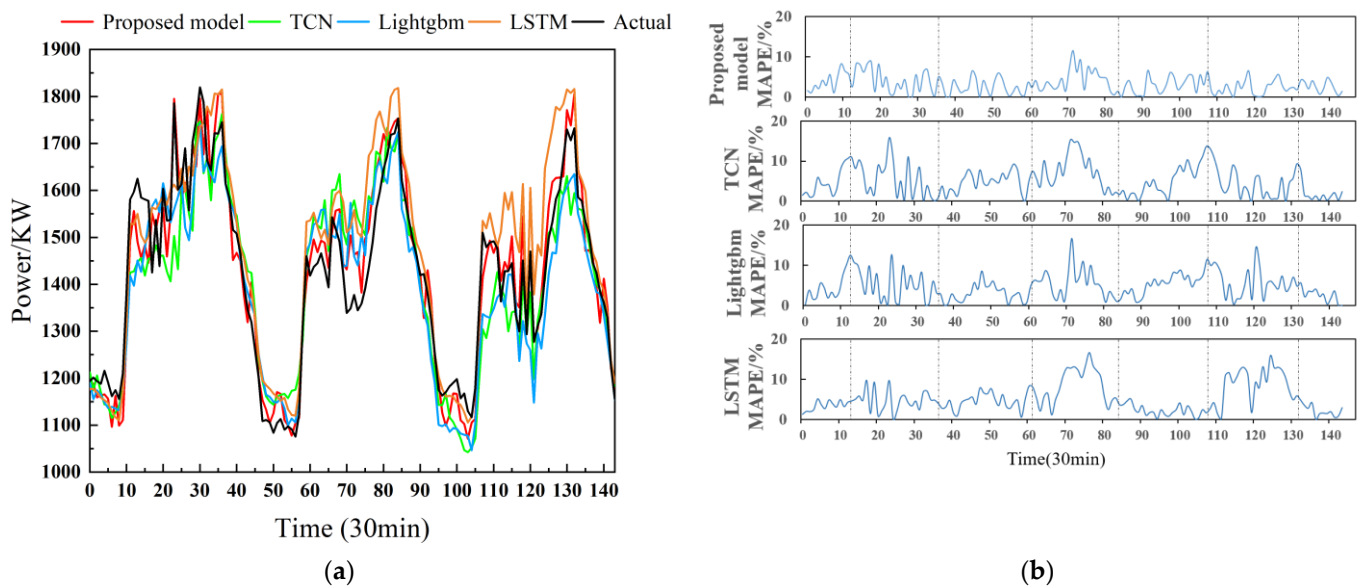


Figure 10. Comparison of multimodel prediction results for scenario 1 (1.1% penetration): (a) Comparison of multimodel prediction curves; (b) comparison of multimodel prediction errors.

From Figure 11a, it can be seen that the power curve shows an obvious “duck shape”, and the peak-to-valley difference is large. Since most of the PV generation is concentrated during midday, a new valley appeared on a clear day. The peak-to-valley difference and the time of peak-to-valley occurrence are determined by the combination of solar irradiance and power consumption features. Figure 11b shows the relative error distribution of each model, which makes it easy to see that the error distribution of the proposed model is more uniform compared to the three comparison models.

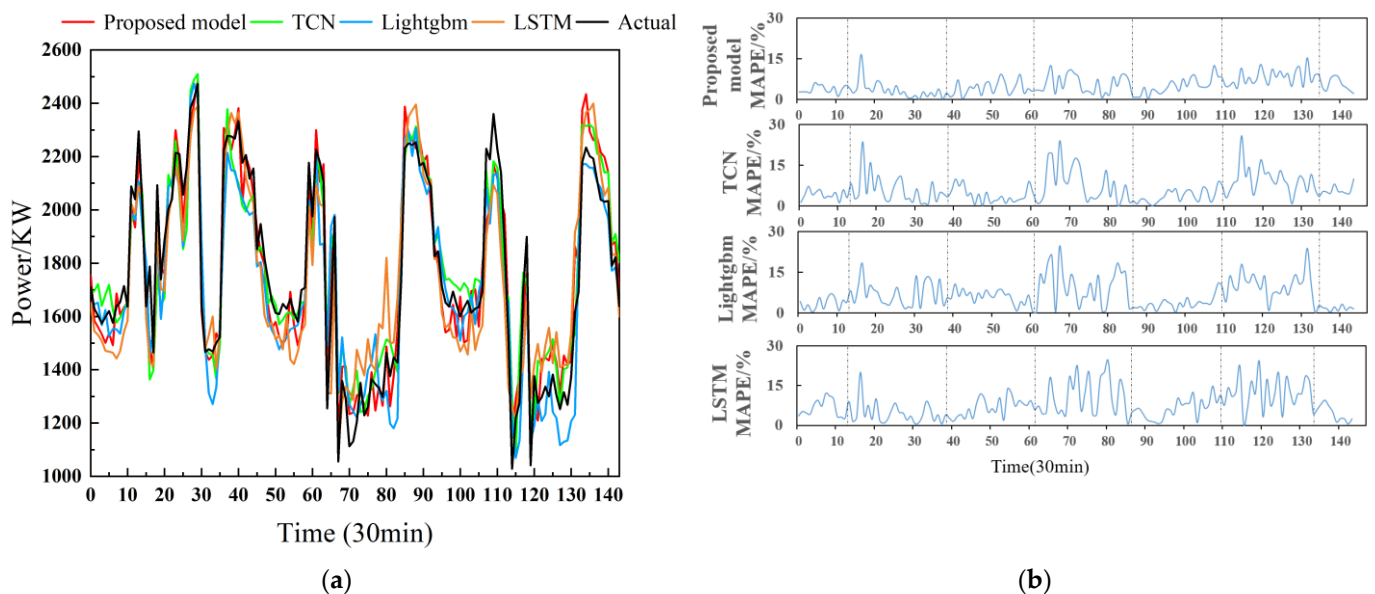


Figure 11. Comparison of multimodel prediction results for scenario 3 (21.7% penetration). (a) Comparison of multimodel prediction curves; (b) comparison of multimodel prediction errors.

In order to fully verify the learning ability of the proposed model regarding the source-load coupling features, we selected the Tempe campus with a 13% penetration as the research object. We simulate two new penetration scenarios by changing the DRE output without changing the electricity load 1 to 30 December 2019, was selected as the simulation period. The results of the measuring metrics are listed in Table 7.

Table 7. Comparison of the measuring metrics for different DRE penetration simulation scenarios.

Model	Scenario 4 (5% Penetration)			Scenario 5 (30% Penetration)		
	MAPE (%)	MBE (KW)	NRMSE	MAPE (%)	MBE (KW)	NRMSE
Proposed model	4.48	−136.23	0.03412	6.98	−362.16	0.04631
TCN	5.32	−224.04	0.03561	9.63	−419.34	0.05721
LSTM	5.62	−235.78	0.03705	10.73	−695.06	0.06031
Lightgbm	5.48	−180.61	0.03623	10.34	−575.93	0.05972

Table 7 shows that the increase in DRE penetration will reduce the prediction accuracy under the same electricity load level, while the proposed model with trend features shows better adaptability than individual models.

To summarize, the proposed model enhances the ability to explore the long-term macroscopic trend, short-term local variations, and time-sharing consumption features, which leads to a significant improvement in the forecasting effect. Intuitively, the proposed model has a higher prediction accuracy and better generalization in different DRE penetration scenarios.

7. Conclusions

This paper proposes a short-term power prediction method based on dynamic segmented curve matching and trend feature perception to improve the accuracy of power prediction. The main conclusions obtained are as follows:

1. We propose a prediction model that takes massive power data, multitime-scale meteorological data, and power feature indexes as the inputs. Through power curve clustering, dynamic segmentation, and trend feature perception, the proposed model can learn the time-sharing consumption features and trend features of power curves to identify the effective information of temporal power features. When compared with other prediction models, the prediction results show that the proposed model is suitable for power prediction with multiple sources of influencing factors and has a higher prediction accuracy;
2. There are different source–load coupling features in different DRE penetration scenarios. In order to fully consider the coupling features, it is necessary to analyze the dominant factors in different DRE penetration scenarios.

Subsequent research will use uncertainty prediction methods for high DRE penetration campuses to improve the model's adaptability.

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

Funding: This work was supported by the Guangxi Special for Innovation-Driven Development (grant numbers AA19254034) and the Guangdong Basic and Applied Basic Research Foundation (Guangdong-Guangxi Joint Foundation) [grant numbers 2021A1515410009].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in the manuscript are downloaded from public access and are open source data. Load data downloaded from <http://cm.asu.edu/> (accessed on 1 June 2022). Weather data downloaded from <https://nsrdb.nrel.gov/> (accessed on 1 June 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Facchini, A. Distributed energy resources: Planning for the future. *Nat. Energy* **2017**, *2*, 17129. [CrossRef]
2. Lotero, R.; de Souza, H. Optimal Selection of Photovoltaic Generation for a Community of Electricity Prosumers. *IEEE Lat. Am. Trans.* **2020**, *18*, 791–799. [CrossRef]
3. Chen, L.; Liu, N.; Yu, S.; Xu, Y. A Stochastic Game Approach for Distributed Voltage Regulation Among Autonomous PV Prosumers. *IEEE Trans. Power Syst.* **2022**, *37*, 776–787. [CrossRef]
4. Hong, T.; Fan, S. Probabilistic electric load forecasting: A tutorial review. *Int. J. Forecast.* **2016**, *32*, 914–938. [CrossRef]
5. Li, K.; Wang, F.; Mi, Z.; Fotuhi-Firuzabad, M.; Duić, N.; Wang, T. Capacity and output power estimation approach of individual behind-the-meter distributed photovoltaic system for demand response baseline estimation. *Appl. Energy* **2019**, *253*, 113595. [CrossRef]
6. Kara, E.C.; Roberts, C.M.; Tabone, M.; Alvarez, L.; Callaway, D.S.; Stewart, E.M. Disaggregating solar generation from feeder-level measurements. *Sustain. Energy Grids Netw.* **2018**, *13*, 112–121. [CrossRef]
7. Charytoniuk, W.; Chen, M.S. Nonparametric regression based short-term load forecasting. *IEEE Trans. Power Syst.* **1998**, *13*, 725–730. [CrossRef]
8. Wang, R.; Zhang, W.; Deng, W.; Zhang, R.; Zhang, X. Study on Prediction of Energy Conservation and Carbon Reduction in Universities Based on Exponential Smoothing. *Sustainability* **2022**, *14*, 1903. [CrossRef]
9. Lee, C.-M.; Ko, C.-N. Short-term load forecasting using lifting scheme and ARIMA models. *Expert Syst. Appl.* **2011**, *38*, 5902–5911. [CrossRef]
10. Li, G.; Li, Y.; Roozitalab, F. Midterm Load Forecasting: A Multistep Approach Based on Phase Space Reconstruction and Support Vector Machine. *IEEE Syst. J.* **2020**, *14*, 4967–4977. [CrossRef]
11. Amjady, N. Short-Term Bus Load Forecasting of Power Systems by a New Hybrid Method. *IEEE Trans. Power Syst.* **2007**, *22*, 333–341. [CrossRef]
12. Shen, Y.; Zhang, J. Short-term load forecasting of power system based on similar day method and PSO-DBN. In Proceedings of the 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2), Beijing, China, 20–22 October 2018. [CrossRef]
13. Inteha, A.; Nahid Al, M.; Hussain, F.; Khan, I.A. A Data Driven Approach for Day Ahead Short-Term Load Forecasting. *IEEE Access* **2022**, *10*, 84227–84243. [CrossRef]
14. Wang, Z.; Jia, L. Short-Term Photovoltaic Power Generation Prediction Based on LightGBM-LSTM Model. In Proceedings of the 2020 5th International Conference on Power and Renewable Energy (ICPRE), Shanghai, China, 12–14 September 2020; pp. 543–547. [CrossRef]
15. Alipour, M.; Aghaei, J.; Norouzi, M.; Niknam, T.; Hashemi, S.; Lehtonen, M. A novel electrical net-load forecasting model based on deep neural networks and wavelet transform integration. *Energy* **2020**, *205*, 118106. [CrossRef]
16. Sun, M.; Zhang, T.; Wang, Y.; Strbac, G.; Kang, C. Using Bayesian Deep Learning to Capture Uncertainty for Residential Net Load Forecasting. *IEEE Trans. Power Syst.* **2020**, *35*, 188–201. [CrossRef]
17. Wang, H.; Zhang, Z. TATCN: Time Series Prediction Model Based on Time Attention Mechanism and TCN. In Proceedings of the 2022 IEEE 2nd International Conference on Computer Communication and Artificial Intelligence (CCAI), Beijing, China, 6–8 May 2022; pp. 26–31.
18. Kaur, A.; Nonnenmacher, L.; Coimbra, C.F.M. Net load forecasting for high renewable energy penetration grids. *Energy* **2016**, *114*, 1073–1084. [CrossRef]
19. Wan, Z.; Xuan, L.; Xiang, R.; Zhang, Y.; Zhang, H. Study on the Electricity Transaction of DG Prosumer under Different Settlement Modes. In Proceedings of the 2021 China International Conference on Electricity Distribution (CICED), Shanghai, China, 6–9 April 2021; pp. 1006–1009.
20. Campus Metabolism. 2022. Available online: <http://cm.asu.edu/> (accessed on 1 June 2022).
21. Song, J.; He, C.; Li, Z. Daily Load Curve Clustering Method Based on Feature Index Dimension Reduction and Entropy Weight Method. *Autom. Electr. Power Syst.* **2019**, *43*, 65–72. [CrossRef]
22. Tambunan, H.B.; Barus, D.H.; Hartono, J.; Alam, A.S.; Nugraha, D.A.; Usman, H.H.H. Electrical Peak Load Clustering Analysis Using K-Means Algorithm and Silhouette Coefficient. In Proceedings of the 2020 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP), Bandung, Indonesia, 23–24 September 2020; pp. 258–262. [CrossRef]
23. Zhang, G.; Zhu, S.; Bai, X. Federated Learning-Based Multi-Energy Load Forecasting Method Using CNN-Attention-LSTM Model. *Sustainability* **2022**, *14*, 2843. [CrossRef]
24. Chen, X.; Chen, W.; Dinavahi, V.; Liu, Y.; Feng, J. Short-Term Load Forecasting and Associated Weather Variables Prediction Using ResNet-LSTM Based Deep Learning. *IEEE Access* **2023**, *11*, 5393–5405. [CrossRef]
25. Acquah, M.A.; Jin, Y.; Oh, B.-C.; Son, Y.-G.; Kim, S.-Y. Spatiotemporal Sequence-to-Sequence Clustering for Electric Load Forecasting. *IEEE Access* **2023**, *11*, 5850–5863. [CrossRef]
26. Dunwen, S.; Xuetao, Y.; Zheng, L.; Hanyang, D. A Transient Voltage Stability Evaluation Model Based on Morphological Similarity Distance Online Calculation. In Proceedings of the 2020 IEEE 3rd Student Conference on Electrical Machines and Systems (SCEMS), Jinan, China, 4–6 December 2020; pp. 243–247.

27. Lai, G.; Chang, W.-C.; Yang, Y.; Liu, H. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor, MI, USA, 6–8 July 2018; pp. 95–104.
28. National Solar Radiation Database (NSRDB). 2022. Available online: <http://nsrdb.nrel.gov/> (accessed on 1 June 2022).
29. Feng, C.; Cui, M.; Hodge, B.-M.; Zhang, J. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. *Appl. Energy* **2017**, *190*, 1245–1257. [[CrossRef](#)]
30. Wang, Y.; Zhang, N.; Chen, Q.; Kirschen, D.S.; Li, P.; Xia, Q. Data-Driven Probabilistic Net Load Forecasting with High Penetration of Behind-the-Meter PV. *IEEE Trans. Power Syst.* **2018**, *33*, 3255–3264. [[CrossRef](#)]

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