Medium-Long Term Electricity Load Forecasting Based on NSNP Systems and Attention Mechanism

Lin Guo
Xihua University
Jun Wang
wj.xhu@hotmail.com

Xihua University
Hong Peng
Xihua University
Yunzhu Gao
Xihua University

Research Article

Keywords: NSNP, Attention, LSTMSNP, Load Forecast

Posted Date: December 11th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3719302/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Journal of Membrane Computing on March 6th, 2024. See the published version at https://doi.org/10.1007/s41965-024-00138-z.

Medium-Long Term Electricity Load Forecasting Based on NSNP Systems and Attention Mechanism

Lin Guo[1,2], Jun Wang[1,2*†], Hong Peng[2,3†], Yunzhu Gao[1,2]

[1*]School of Electrical Engineering and Electronic Information, Xihua University, No.999 Jinzhou Road, Chengdu, 610039, Sichuan, China.
[2]Sichuan Province Key Laboratory of Power Electronics Energy-saving Technologies & Equipment, Xihua University, No.999 Jinzhou Road, Chengdu, 610039, Sichuan, China.
[3]School of Computer and Software Engineering, Xihua University, No.999 Jinzhou Road, Chengdu, 610039, Sichuan, China.

*Corresponding author(s). E-mail(s): wj.xhu@hotmail.com;
Contributing authors: guolin_ll@stu.xhu.edu.cn; ph.xhu@hotmail.com; gaoyunzhu@stu.xhu.edu.cn;
†These authors contributed equally to this work.

Abstract

Accurate load forecasting can provide important information support for intelligent operation of power systems, it can assist the power grid to deploy production plans in advance to uphold the equilibrium between the supply and demand for electrical power, or plan investment strategies based on the results of the forecast. Nonlinear Spiking Neural P (NSNP) system [1] belongs to a category of computational systems with distributed, parallel and non-deterministic characteristics that have the analytical skill to solve nonlinear problems. Aiming at the temporal characteristics and complex nonlinear characteristics of electrical load data, this paper proposes a new Medium-Long Term Load Forecast model LF-ASNP based on NSNP system and attention mechanism, which can accurately analyze the characteristics of historical load data and forecast the electrical load. In this paper, the LF-ASNP model is validated in several benchmark datasets, and the analysis of the experimental results fully demonstrates that the model can forecast the power load effectively and reliably.

Keywords: NSNP, Attention, LSTMSNP, Load Forecast

1 Introduction

One of the important foundations of power system security and stability is effective and accurate load forecasting. Effective load forecasting is the basis of smart grid services, which can assist in production scheduling, demand response, price adjustment, risk control and development planning of the electrical power system, and can also improve the security, reliability and economy of power system operation.

In real production and life, different load forecasting tasks have different needs, so the forecasting time lengths set for different tasks are different, and different models have different definitions for the length of time. For the classification of forecasting time length is roughly shown in Fig.1.

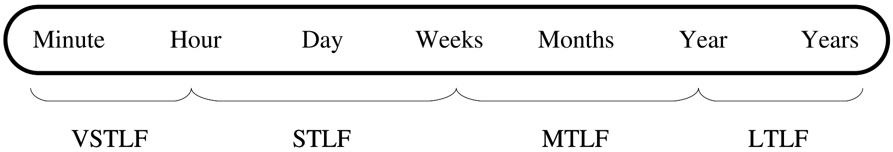

Fig. 1 Classification of load forecasts based on length of time

Load forecasting can be broadly categorized into four types based on different forecasting time lengths [2]: very short term load forecasting (VSTLF), short term load forecasting (STLF), medium term load forecasting (MTLF), and long term load forecasting (LTLF). Forecasts of different time length types excel at solving different problems. For example, VSTLF [3] can provide information support for real-time power market operation, adjust the electricity price according to the forecast results to indirectly affect the distribution of loads; it can also help the power grid to realize online scheduling and reduce the cost of power generation. STLF [4, 5] can standardize and guide the daily management of electric power companies, arrange a reasonable power consumption plan, strengthen the utilization efficiency of the grid, and serve as the information basis for the daily operation of the power grid. MTLF [6] is suitable for overhaul of production equipment, planning of production and operation, power scheduling, and making long-time operation plans. LTLF [7] can be used for planning, capacity increase and reconstruction of the power grid, such as addition of new electrical facilities to the power system, and the switching of large power facilities, etc., and its prediction results can provide a reference standard of power consumption for a long period of time. Compared with VSTLF and STLF, MTLF and LTLF are affected by more factors, such as population change, urban economic development, changes in the natural environment, changes in the international energy market, etc., so their forecasts are more difficult.

In load forecasting, researchers had proposed various types of load forecasting methods such as trend extrapolation, linear regression or nonparametric regression, support vector regression (SVR), Kalman Filtering, etc.. e.g., Reference [8] used decomposition and bootstrap aggregation methods and Auto-regressive Moving Average (ARIMA) to do univariate forecasting. Reference [7] used Particle Swarm Optimization (PSO) and data from Egyptian and Kuwaiti grids to forecast annual peak loads. Reference [9] predicts loads by combining different regression models (bagging, random forest, extra trees, ada boost, and gradient boosting regressors). Reference [10] utilized expert forecasting method and fuzzy Bayesian model to predict load per capita. Based on the gray system theory in Reference [11], variants of gray system theory models [5, 12, 13] have been proposed to make predictions of load. Reference [6] utilizes three regression techniques: linear, compound growth, and quadratic regression to accomplish the predictions. The approach used in the Reference [14] is exponential smoothing (ES). These types of traditional methods are easy to implement, but they are less capable of extracting the increasing nonlinear features of the grid nowadays, and thus less accurate and flexible.

Artificial neural network models have shown enough capability in solving nonlinear problems and hence are widely used in various fields[2]. Many neural network models are now proposed, such as Multi-layer Perceptron (MLP)[15], Radial Basis Function Network (RBFN)[16], Wavelet Neural Network (WNN)[17], Support Vector Machine (SVM) [18], Extreme Learning Machine (ELM)[17], Fuzzy Neural Network (FNN)[19], and so on. The main drawback of shallow neural networks is the potential "overfitting". As relevant research advances, this problem has been effectively addressed and new models are constantly being proposed, such as Recurrent Network (RNN), Gate Recurrent Network (GRU), Deep Residual Network (DRN), Convolutional Neural Network (CNN)[20], and Long Short-Term Memory (LSTM) [21–23], and Temporal Convolutional Network (TCN)[24]. New deep neural networks have also been proposed and applied to power system load forecasting in recent years. Reference [25] presented cycle-based long short-term memory (C-LSTM) and time-dependent convolutional neural network (TD-CNN) to predict loads of different lengths. In Reference [26], an Artificial Neural Network (ANN), Linear Regression (LR) and Adaptive Augmentation models were used to predict loads of different lengths. Reference [27] combined ARIMA and ANN thus obtaining the OSA-LP model and made load forecasts for local countryside. Reference [28] proposed an electricity load peak building forecasting method based on k-means clustering and auto-ARIMA, and verified the validity of the model by using the load dataset of Chubu University (East Campus) for the year 2017-2018. Reference [3] proposed new forecasting methods based on Bayesian regularization (BR) and Levenberg-Marquardt (LM) for load forecasting of buildings. The method used in Reference [29] is an improved Elman neural network (IENN) and novel shark smell optimization (NSSO) algorithm. Reference [30] proposed a hybrid prediction method by combining CNN and LSTM-AE (autoencoder). Reference [31] used Multicolumn radial basis neural network (MCRN) and Radial base function neural network (RBFN) for load forecasting. Reference [32] did prediction experiments with sequential-grid-approach (SGA) based SVR model on two benchmark datasets.

Membrane computing constitutes a category of computational models abstracted from the interaction mechanisms between biological cells. The Spiking Neural P

(SNP) system [33] inspired by spiking neurons belongs to the third generation of neural network that are capable of modeling the interaction mechanism of spiking signals between cells, and it was proved Turing-complete. The NSNP system[1] is a nonlinear variant of the SNP system. NSNP is a new type of neural network characterized by distributed, parallelism and bio-like flexibility. Compared to the SNP system, in the NSNP system, each neuron contains an internal state and a nonlinear spiking rule, and the neuron's state is computed using the firing rule. Thus, the NSNP system has the ability to solve the nonlinear power system load forecasting task. In this study, a new recursive type model LF-ASNP is proposed based on NSNP system to solve the engineering problem of power load forecasting.

The subsequent sections of the paper are structured as follows. The LF-ASNP model is described in Section 2. Section 3 contains the experimental setup and results. And Section 4 concludes the study.

2 Prediction Model

2.1 NSNP System

The Spiking Neural P (SNP) system is a new model of membrane computation drew inspiration from mechanism of firing interactions between biological neuronal cells. The NSNP system, on the other hand, represents a nonlinear variation of the former, which makes NSNP different from SNP and traditional membrane computation models by focusing more on the analysis of the characteristics of nonlinearity. And it can be defined as follows:

$$\Pi = (O, \sigma_1, \sigma_2, \cdots, \sigma_k, synapses, x, y) \tag{1}$$

Where

1. $O = a$ is a singleton alphabet. The a denotes a spike.
2. $\sigma_i = (u_i, r_i)$ represents the ith cell, $i = 1, 2, \cdots\cdots\cdots, m$, and u_i represents the initial value of this neuron. The r_i represents the nonlinear spiking rules contained in σ_i, which are of the form $a^{g(u)} \to a^{f(u)}$. In the firing rules, $u = u_i + x$, and $g(\cdots)$ and $f(\cdots)$ represent nonlinear functions of the number of spikes consumed and generated, respectively.
3. $synapses = 1, 2, \cdots, k \times 1, 2, \cdots, k$, that represents information about connections between cells and $i, j = 1, 2, \cdots, k$, $m \neq n$.
4. x and y represent inputs and outputs from outside, respectively.

Fig.2 shows a simple INPUT module of NSNP. And to make it easier to understand, let b(x)=x, At the moment of $t = 0$, let the neuron σ_{c_1} receives 1 spike from the environment, and the spiking rule of σ_{c_1} is $1|a^x \to a^{\beta(x)}$, at this time, it satisfies the spiking rule, so σ_{c_1} will consume x spike, and generate $b(x)$ spikes and send them to σ_{c_2} and σ_{c_3}, since $b(x) = x$, the number of spike consumed by σ_{c_1} and the number of spike send to σ_{c_2} and σ_{c_3} are both 1. At $t = 2$, σ_{c_2} satisfies the trigger condition and

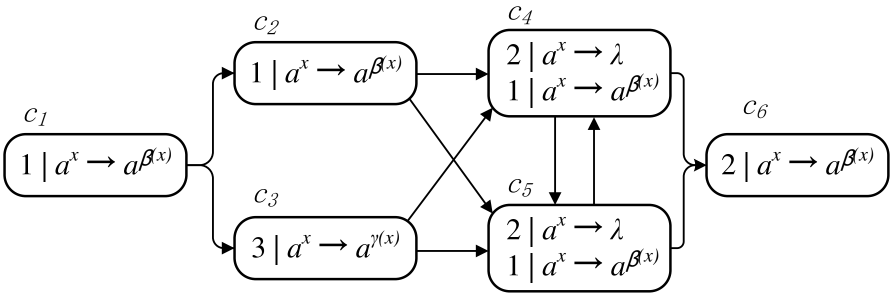

Fig. 2 NSNP INPUT module

consumes 1 spike, and sending 1 spike to σ_{c_4} and σ_{c_5}; meanwhile, σ_{c_3} does not satisfy the trigger condition, so it is not activated. At $t = 3$, σ_{c_4} and σ_{c_5} are triggered at the same time and send 1 spike to each other, so the number of saved spikes in σ_{c_4} and σ_{c_5} is still 1. Moreover, σ_{c_4} and σ_{c_5} send 1 spike to σ_{c_9} at the same time, therefore, at $t = 4$, σ_{c_9} contains 2 spikes internally and can be triggered.
In the NSNP system, each unit contains an internal state u and one or more nonlinear spiking rules, the state is computed based on the spiking rules. When practically applied to engineering tasks, at time step $t + 1$, the neuron of NSNP can compute internal state and output by the following formulas:

$$u_i\left(t+1\right) = u_i\left(t\right) - g\left(u_i\left(t\right) + x_i\left(t+1\right)\right) + n \tag{2}$$
$$y_i(t+1) = f\left(u_i\left(t\right) + x_i\left(t+1\right)\right) \tag{3}$$

2.2 LSTMSNP

The LSTMSNP model [34] is inspired by the NSNP system and the LSTM model. A single cell in the NSNP system is taken to obtain parameterized model with nonlinear firing rules and nonlinear gate functions (reset gate, consumption gate, and generation gate), which in turn yields the recursive model LSTMSNP. its state update equation and status calculation equation are:

$$\mathbf{h}\left(t+1\right) = \mathbf{w}_h f\left(\mathbf{w}_u \mathbf{u}\left(t\right) + \mathbf{w}_x \mathbf{x}\left(t+1\right) + \mathbf{b}\right) \tag{4}$$

$$\mathbf{u}\left(t+1\right) = \mathbf{w}'_u \mathbf{u}\left(t\right) - g\left(\mathbf{w}''_u u\left(t\right) + \mathbf{w}'_x x\left(t+1\right) + \mathbf{b}'\right) \tag{5}$$

Where $\mathbf{w}_h, \mathbf{w}_u, \mathbf{w}_x, \mathbf{w}'_u, \mathbf{w}''_u$ and $\mathbf{w}'_x$ represent the weight values, $\mathbf{b}$ and $\mathbf{b}'$ are the bias values.

Building upon the above principle, a recursive-like neuron model, LSTMSNP, is developed. LSTMSNP contains three nonlinear gates: 'reset gate', 'consumption gate', and 'generation gate', which respectively determine the number of previous state resets $r\left(t\right)$, the number of previous state consumption $c\left(t\right)$ and the number of spikes output $o\left(t\right)$ and their update equations are as follows:

$$\mathbf{r}\left(t+1\right) = \Phi\left(\mathbf{U}_r \mathbf{u}\left(t\right) + \mathbf{W}_r \mathbf{x}\left(t+1\right) + \mathbf{b}_r\right) \tag{6}$$

$$\mathbf{c}(t+1) = \Phi\left(\mathbf{U}_c\mathbf{u}(t) + \mathbf{W}_c\mathbf{x}(t+1) + \mathbf{b}_c\right) \tag{7}$$

$$\mathbf{o}(t+1) = \Phi\left(\mathbf{U}_o\mathbf{u}(t) + \mathbf{W}_o\mathbf{x}(t+1) + \mathbf{b}_o\right) \tag{8}$$

Where $\mathbf{u}(t+1)$ is the value of the cell's input at the moment $t+1$, and $\mathbf{u}(t)$ is the output at (t), $\Phi(\cdot)$ is a nonlinear function. Let $f \equiv g$, and define the number of spikes generated as $a(t)$:

$$\mathbf{a}(t+1) = f\left(\mathbf{U}_a\mathbf{u}(t) + \mathbf{W}_a\mathbf{x}(t+1) + \mathbf{b}_a\right) \tag{9}$$

Then Eq.(4) and Eq.(5) can be transformed into:

$$\mathbf{h}(t+1) = \mathbf{o}(t+1) \odot \mathbf{a}(t+1) \tag{10}$$

$$\mathbf{u}(t+1) = \mathbf{r}(t+1) \odot \mathbf{u}(t) - \mathbf{c}(t+1) \odot \mathbf{a}(t+1) \tag{11}$$

Where $\cdot$ is the inner product of the two vectors, and $\mathbf{W}_r$, $\mathbf{W}_c$, $\mathbf{W}_o$, $\mathbf{W}_a$, $\mathbf{U}_r$, $\mathbf{U}_c$, $\mathbf{U}_o$, $\mathbf{U}_a$, $\mathbf{b}_r$, $\mathbf{b}_c$, $\mathbf{b}_o$, and $\mathbf{b}_a$ in Eq.(6)-Eq.(10) are trainable.

The LSTMSNP is a parameterized NSNP system, therefore it can effectively learn the nonlinear features of the load data, based on which we introduce an attention mechanism to assist in enhancing the performance of the model to capture the temporal features of the load data by quantitatively assigning specific time-step weight values.

2.3 Attention

The attention mechanism simulates the human brain, and reasonably allocate limited attention resources. Similar to a human being processing a large amount of visual data, brain selectively focuses its attention on the regions that need to be focused on, and reduces or even ignores attention to other unimportant regions, so as to obtain more important and effective information. Significant regular features in time-series data tend to contain more detailed information and have a greater impact on the actual trend. The attention mechanism optimizes the efficiency of information processing by applying limited resources to the acquisition of key information.

Specific weights $\mathbf{w}_A$ are assigned by calculating the degree of correlation between different time steps, it quantifies the correlation between inputs at different time steps. In the attention layer, define $\mathbf{x}_A$ as input, and $\mathbf{x}_A$ is equal to $\mathbf{h}$, which is the output of previous layer. The weights are $\mathbf{W}_A$ and n is the time step, $\mathbf{b}_A$ is bias vector, $\mathbf{z}$ is the attention component, and the formula for the attention component is:

$$\mathbf{z} = \tanh\left(\mathbf{x}_A \odot \mathbf{W}_A + \mathbf{b}_A\right) \tag{12}$$

The attention components are normalized with the Softmax function to obtain the ith time step component w_i of attention weights $\mathbf{w}_A$:

$$w_i = \frac{\exp(z_i)}{\sum_{j=1}^{n} \exp(z_j)} \tag{13}$$

Finally the context vector **c** is computed one by one :

$$c_i = \sum_{j=1}^{n} w_j \mathbf{x}_j \tag{14}$$

Where $\mathbf{x}_j$ is the ith time step component of $\mathbf{x}_A$, $\mathbf{W}_A$ and $\mathbf{b}_A$ are trainable.

2.4 Overview

In this study, a new recursive model for load forecasting is constructed, which introduces an Attention mechanism to enhance the model's ability to acquire temporal features and important fine-grained features for power load data.

Fig.3 demonstrates the structure of proposed model and the process of load prediction. We add the attention layer behind the LSTMSNP, and the input of the attention layer is the output of the previous layer. After getting the output of the LSTMSNP layer, the correlation between the outputs of different time steps is calculated so as to assign attention weights to the outputs of the LSTMSNP at different time steps, and the attention weights are weighted and summed with the output of the previous layer, and then the model outputs the results.

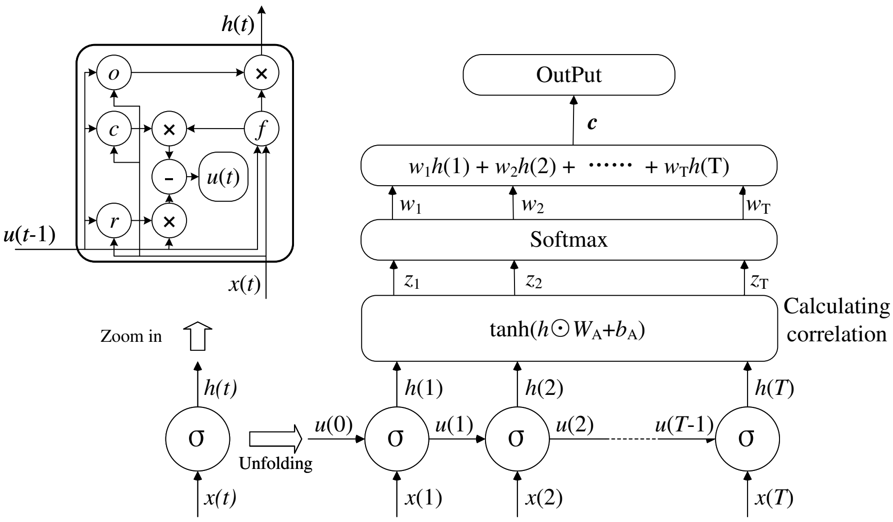

Fig. 3 Structure of LF-ASNP

3 Experiment

3.1 Experimental setting

To implement the model, we used the keras machine learning platform, which encapsulates the tensorflow framework. And LSTMSNP was added in the Recurrent base class of keras for easy calling. The experiments are based on GPU implementation. The graphics card used is NVIDIA-A40. Table.1 shows the parameter configuration used for the experiment.

Table 1 Hyper-Parameter settings

Hyper-Parameter	Value	Hyper-Parameter	Value
Activation	tanh	Recurrent Dropout	0
Recurrent Activation	hard sigmoid	Number of Neurons	24
Kernel Initializer	glorot uniform	Epoch	100
Recurrent Initialize	orthogonal	Loss	MSE
Bias Initializer	zeros	Optimizer	adam
Dropout	0.1	Output Layer	Dense(1)

3.2 Assessment of indicators

To measure the predictive performance of LF-ASNP and compare the results with other benchmark methods, five evaluation metrics were chosen for the experiment: mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), normalized root mean square error (nRMSE), and R-squared (R2).

MAE, which assesses the accuracy of the predictive model by calculating the average of the absolute differences between the prediction and real observations, is formulated as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{15}$$

MAPE, which enables the calculation of the mean of the absolute percentage errors, with the expression:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{16}$$

RMSE, which is widely used as a measure of prediction error, is the square root of the mean of the squared differences between prediction and real observations, expressed as:

$$RMSE = \frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \tag{17}$$

nRMSE, which is a normalized measure of RMSE, eliminating the effect of the scale by dividing by the range of the observations, expressed as

$$nRMSE\% = \frac{RMSE}{(y_{max} - y_{min})} \times 100 \tag{18}$$

R2 Score, a measure of the degree of fit of the regression model, which indicates the proportion of the variance of the actual observations that is explained by the predicted values, expressed as:

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{19}$$

Where the $\hat{y}_i$ is the prediction of the model; y_i is the real value; $\bar{y}$ is the average of the real historical data; n is the length of the predicted sequence. Among the above error indicators, the smaller the values of MAE, MAPE, RMSE, nRMSE, the more accurate and effective the prediction results are; the closer R2 is to 1, the more the prediction results are fitted to the real load data.

3.3 Experiment A

3.3.1 Datasets

The AUS dataset is supplied by Australian Energy Market Operator (AEMO) and contains historical load, temperature and humidity, and electricity price data recorded every half hour from 2006 to 2010 in Australia. Based on the AUS dataset, we set up a set of comparison experiments with the same dataset allocation strategy as the 2 benchmark models [35], using the last 15% of the AUS dataset as the test set.

3.3.2 Results

In Fig.4, the upper and lower parts are the load curve and the absolute error curve, respectively. And it can be seen that the fitting between the prediction curve and the actual hourly load curve is accurate and reliable, and the absolute error curve in the lower part of the graph shows that the distribution of the error is in a stable and small range, and there is no abnormally distributed extreme value point, which play a crucial role in ensuring the electrical system operating steadily. Moreover, after the introduction of Attention, the number of peak extremes of the absolute error is reduced compared with the previous one, and the error is further reduced, which indicates that LF-ASNP can learn the characteristics and trends of the load data well.

Table.2 shows the comparison of the error indicators of the prediction results. Before the introduction of the attention mechanism, the prediction performance of LSTMSNP is slightly lower than the predictions of the benchmark models ANN and Regression Trees [35], while the parameters of the LF-ASNP model after the introduction of the attention mechanism are significantly improved, the RMSE is reduced by 42.534% compared to the previous one, the nRMSE is reduced by 46.043%, the MAE is reduced by 43.323%, the MAPE is reduced by 44.494% and the R2 is improved by 0.0199. The MAE is reduced by 43.323%, MAPE is reduced by 44.494%, and R2 is improved by 0.0199. It can be clearly seen that the predictive performance is significantly improved after the introduction of the attention mechanism, and it is much better than the ANN and Regression Trees. Fig.5 shows the experimental results in shorter time period, which provides a more intuitive look of the predictions, and in the lower half of the figure it can also be seen that the errors of LF-ASNP are significantly and generally lower than those of the LSTMSNP model. These results demonstrate which proves the superiority of our model.

Table 2 Comparison of Error Indicators of Load Forecasting Results for Australian Dataset

Models	RMSE	nRMSE	MAE	MAPE	R2
ANN	-	-	167.91	1.90	-
Regression Trees	-	-	136.39	1.98	-
LSTMSNP	226.493	3.045	172.283	2.016	0.9705
LF-ASNP	130.156	1.643	97.644	1.119	0.9904

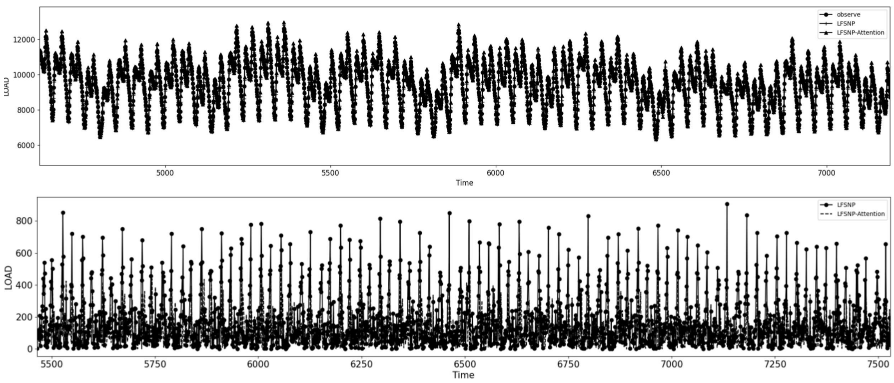

Fig. 4 Australian load curve (top) and absolute error curve (bottom)

3.4 Experiment B

3.4.1 Datasets

The PJM dataset is published by PJM Interconnection, and we chose the same dataset as the baseline model [36], which overall contains historical hourly load data for U.S. regions from 2008 to 2016. Based on the PJM dataset, 2 sets of experiments were set up using the same dataset allocation strategy as the benchmark model. The first set of experiments selects the historical hourly load data from 2008-2014 to train the model, the historical load of 2015 is selected to test model and compare with the 2 benchmark models [36]; the second set of experiments selects the historical hourly load data from 2013-2015 to train the model, and the historical load of 2016 is selected as the test dataset to compare with the 7 benchmark models [36].

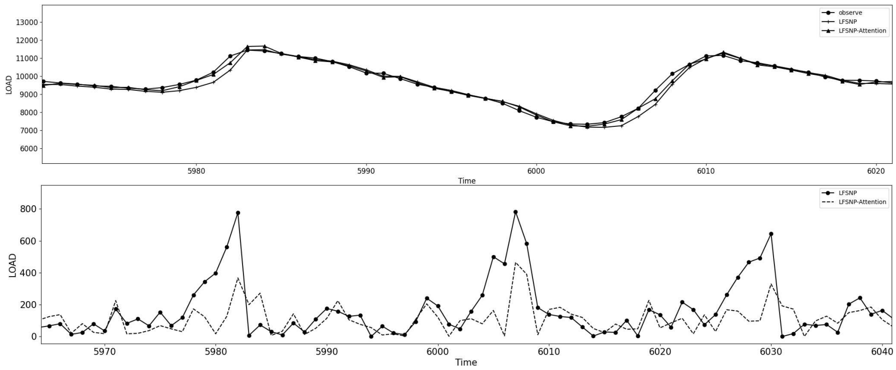

Fig. 5 A portion of the Australian load forecast results

3.4.2 Results

The upper and lower parts in Fig.6 are the load curves and the absolute error curves of the first set of experiments, respectively. The load prediction curves are well fitted to the real load, and the absolute error curves are stably distributed in a small range. Fig.7 Zooming in on the experimental results, it can be observed that the errors of LF-ASNP are generally lower than those of the benchmark model, especially at the extreme value points, the prediction error extreme point of the LF-ASNP model is slightly lower than LSTMSNP, which once again proves the feasibility of the introduced mechanism.

Table.3 is the comparative analysis of the error indicators of LF-ASNP, MTS ARMA and LSTMSNP models in the first set of experiments. The LSTMSNP model outperforms the MTS ARMA, and its nRMSE and MAE are slightly better than the LF-ASNP, whereas the LF-ASNP improves in the MAPE and R2 parameters in comparison with the LSTMSNP. The upper and lower parts of Fig.8 show the predicted load curves and absolute error curves of the second set of experiments, respectively. The fitting effect of the load curves also performs well, and the absolute error curves are also stabilized in a low range, which proves the stability of the model prediction performance. Fig.9 zooms in on the experimental results, and in the upper part of the figure, it can be observed that the error of the LF-ASNP is smaller near the peaks, valleys, and load mutation points, and in the lower part of the figure, it can be similarly observed that the error of the proposed model is smaller at the point of great magnitude. The error indicators of the LF-ASNP model compared to the seven benchmark models (MMPF, AICC, MLPNN, RBFNN, PCR, PLSR, LSTMSNP) for the second set of experiments are shown in Table.4 The data shows that our model has the superior predictions among the mentioned benchmark models, in comparison to the second LSTMSNP, our model demonstrates improvements, with a 1.68% reduction in RMSE, a 3.49% reduction in nRMSE, a 3.88% reduction in MAE, a 4.65% reduction in MAPE, and a 0.001 improvement in R2.

It is proved that the introduced Attention layer can improve the prediction performance of the model, and the various error indicators have been relatively optimized. From Fig.6 and Fig.8, it can also be observed that the Maximum value of the absolute error is mainly distributed in the irregular peak moments in the daily load, and such irregular peaks are mostly caused by the compounding of multiple reasons, Because the PJM dataset covers the total load of the various types of regional integrated loads. Therefore, when some events that have a great impact on certain regions, such as the cold wave, changes in the economic policy, etc., a large number of loads are affected by this factor and the load fluctuates more, and the changes are displayed on the curve of total loads.
From Fig.7 and Fig.9, we can see that the prediction errors of our model are smoother and more stable near the load mutation time point and the point of the extreme value of the error, showing a stable prediction performance.

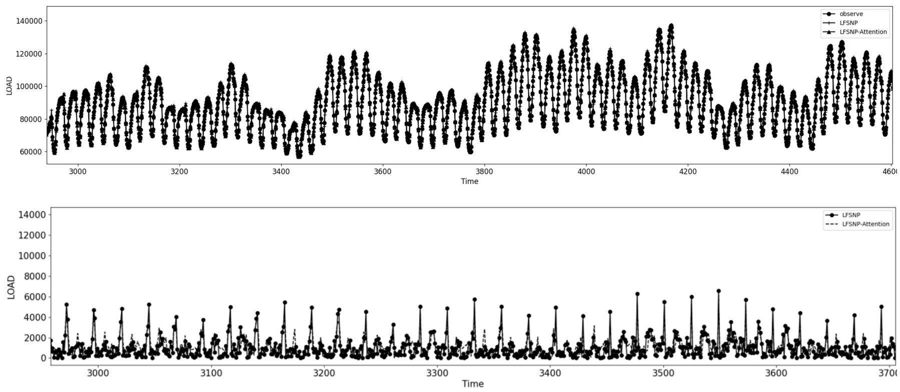

Fig. 6 Curve fit (top) and error curve (bottom) of 2015 load predictions for the U.S. PJM market

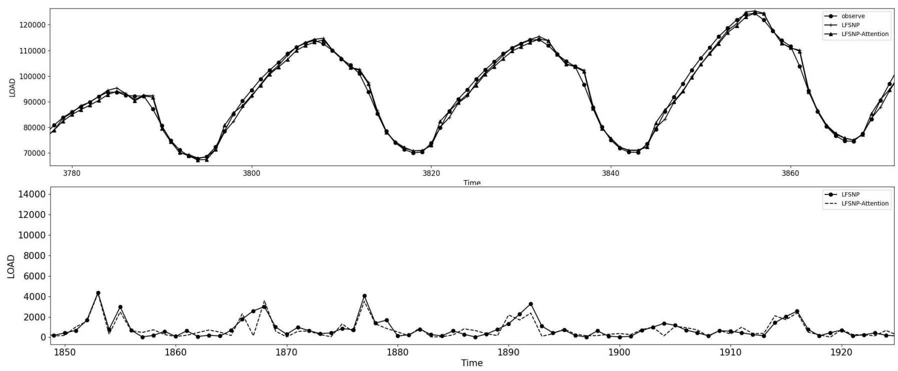

Fig. 7 A portion of 2015 load predictions for the U.S. PJM market

Table 3 Comparison of Error Indicators for 2015 Load Forecast Results for U.S. PJM Markets

Models	Year	Location	nRMSE	MAE	MAPE	R2
MTS ARMA	2015	PJM Network	-	4.68	4.4	-
LSTMSNP	2015	PJM Network	1.6	1.12	1.24	0.9923
LF-ASNP	2015	PJM Network	1.7	1.16	1.15	0.9924

Note: The same formula for MAE, MAPE for dataset B was chosen as in Reference [36].

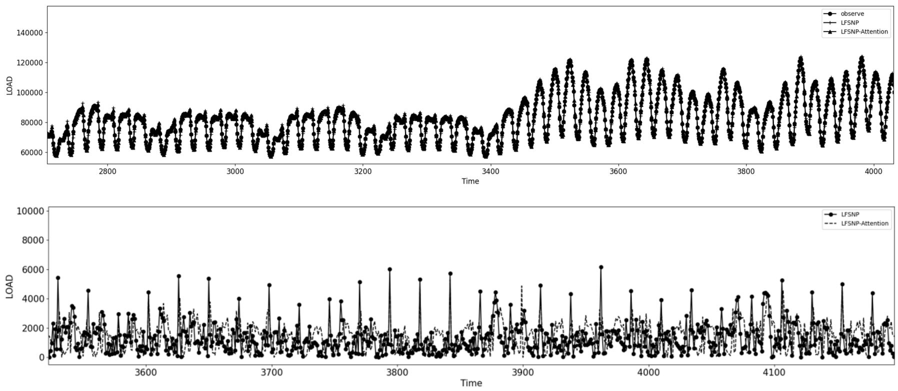

Fig. 8 Curve Fit (Partial) (Top) and Error Curve (Bottom) for 2016 Load Forecast Results for U.S. PJM Markets

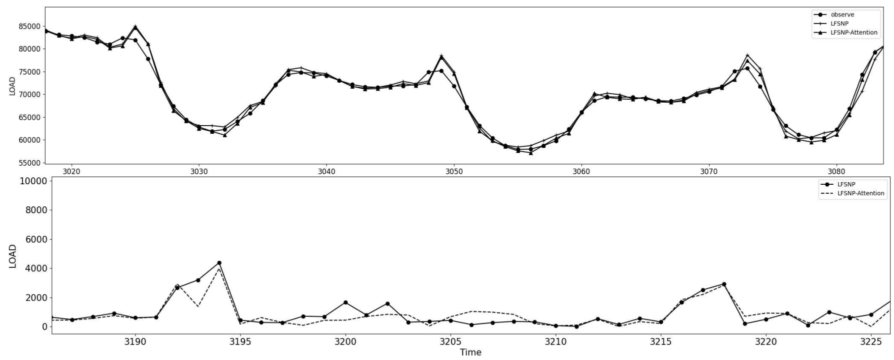

Fig. 9 A portion of 2016 load forecast results for the U.S. PJM market

3.5 Experiment C

3.5.1 Datasets

The GEFCOM2014 dataset is a publicly available dataset used in the Global Energy Forecasting Competition 2014 and contains data on loads, electricity prices, and climate. Following the dataset strategy of the benchmark models [37], only hourly loads from 2005 to 2011 were selected for the experiments. Based on this dataset, one set of

Table 4 Comparison of Error Indicators for 2016 Load Forecast Results for U.S. PJM Markets

Models	Year	Location	RMSE	nRMSE%	MAE	MAPE%	R2
MMPF	2010	Greece	-	-	-	1.87	-
AICC	2010	Greece	-	-	-	1.98	-
WNN	2014	Alberta Canada	1.328	-	0.983	-	-
MLPNN	2014	Alberta Canada	3.724	-	2.556	-	-
RBFNN	2014	Alberta Canada	2.287	-	1.712	-	-
PCR	2016	Polish Network	-	-	-	1.15	-
PLSR	2016	Polish Network	-	-	-	1.09	-
LSTMSNP	2016	PJM Network	1.79	1.72	1.29	1.29	0.992
LF-ASNP	2016	PJM Network	1.76	1.66	1.24	1.23	0.993

Note: The same formula for MAE, MAPE for dataset B was chosen as in Reference [36].

comparison experiments with six benchmark models was set up. Same as the benchmark model, all load data except the last 12,120 load data were selected to train the model.

3.5.2 Results

The load curves in the upper half of Fig.10 show that the predicted load curves can effectively fit the actual load curves. The absolute error curves in the lower half are also stably distributed in a small range, and the number and range of the extreme value points of the error curves are further optimized after the introduction of Attention. In the upper half of Fig.11, it demonstrates that the predicted load curve of the LF-ASNP fits the true load curve better than the baseline model; in the lower half it is also observed that the error distribution of the LF-ASNP model is lower and more stable. Table.5 presents the comparison of the error indicators of the LF-ASNP model with six benchmark models (Exponential Smoothing, ARIMA, Random Forests, RNN, S2S-RNN, LSTMSNP), and the comparison show that the predictions of LF-ASNP are the best in all the five error indicators when the benchmark model [37]. Compared to the second LSTMSNP model, LF-ASNP reduces the RMSE by 2.21%, nRMSE by 6.38%, MAE by 6.91%, MAPE by 12.64%, and the R2 improvement by 0.006.

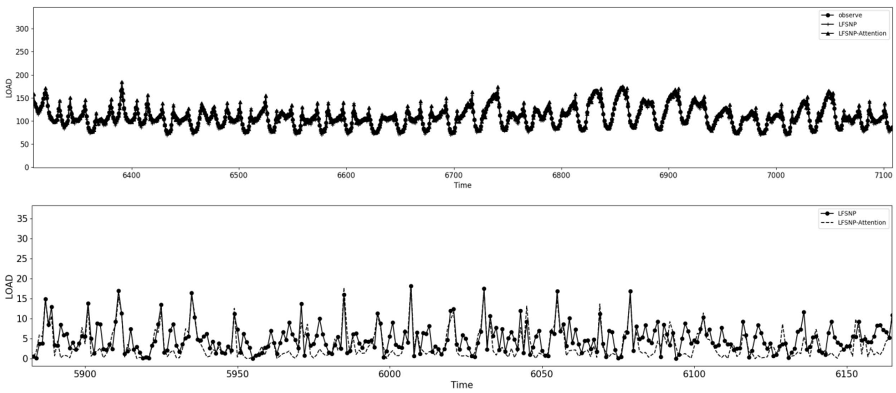

Fig. 10 GEFCOM2014 load curve (top) and error curve (bottom)

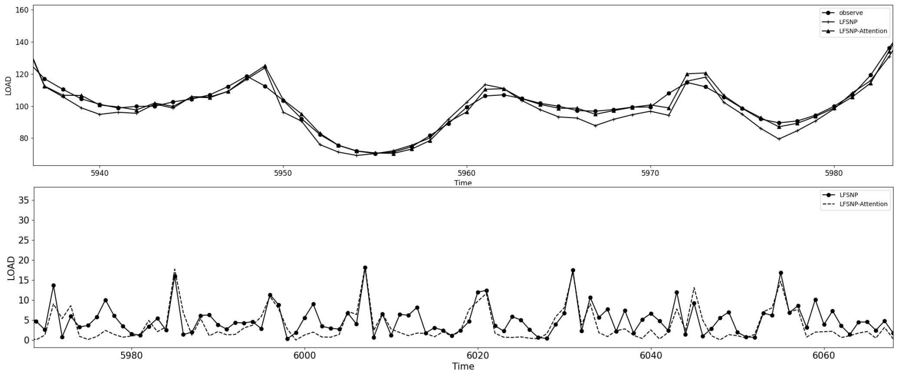

Fig. 11 A portion of load forecast results for GEFCOM2014

Table 5 Comparison of Error Indicators for GEFCOM2014 Load Forecast Resultss

Models	RMSE	nRMSE%	MAE	MAPE%	R2
Exponential Smoothing	24.12	8.00	-	-	0.6245
ARIMA	18.72	6.21	-	-	0.7553
Random Forests	18.31	6.07	-	-	0.8335
RNN	14.37	4.77	-	-	0.9168
S2S-RNN	10.82	3.59	-	-	0.9497
LSTMSNP	5.620	1.865	4.182	2.966	0.9881
LF-ASNP	5.496	1.746	3.893	2.591	0.9887

4 Conclusion

In this study, we propose a recursive model LF-ASNP based on LSTMSNP and Attention mechanism, which can be used for dynamic system modeling due to its nonlinear spiking mechanism. The LSTMSNP model has already possessed the basic load prediction capability, while the Attention layer can effectively improve the distraction problem of LSTMSNP by adjusting the the contribution of different features to the output, so that the model can better capture the time-series characteristics of historical loads, and enhance the performance of model. In the experiments of this paper, prediction error of our model is lower and smoother, the error distribution is stabilized within a range, and it can fit the real load curve well even when the load changes abruptly, and the prediction performance is more stable, the LF-ASNP model can provide reliable load forecasting data for power systems. The results in the three benchmark datasets prove that it can effectively learn the characteristics of load data, and the comparisons prove that the LF-ASNP has better performance. The introduction of the attentional mechanism significantly improves the model's ability to analyze the time-series information, so that it can better achieve the load forecasting task, which can effectively help the power system avoid potential load risks, formulate countermeasures in advance, and guarantee the secure and enhance the robustness of the power system.

Data Availability

The data employed in the experiments is provided within this article.

Conflict of Interest

All authors disclosed no relevant relationships.

Funding Statement

This work was supported by a grant from Chengdu science and Technology Bureau (No. 2023-JB00-00002-SN)

References

[1] Peng, H., Lv, Z., Li, B., Luo, X., Wang, J., Song, X., Wang, T., Pérez-Jiménez, M.J., Riscos-Núñez, A.: Nonlinear spiking neural p systems. International Journal of Neural Systems **30**(10), 2050008 (2020)

[2] Azeem, A., Ismail, I., Jameel, S.M., Harindran, V.R.: Electrical load forecasting models for different generation modalities: A review. IEEE Access **9**, 142239–142263 (2021) https://doi.org/10.1109/ACCESS.2021.3120731

[3] Dagdougui, H., Bagheri, F., Le, H., Dessaint, L.: Neural network model for short-term and very-short-term load forecasting in district buildings. Energy and Buildings **203**, 109408 (2019)

[4] Zhang, W., Chen, Q., Yan, J., Zhang, S., Xu, J.: A novel asynchronous deep reinforcement learning model with adaptive early forecasting method and reward incentive mechanism for short-term load forecasting. Energy **236**, 121492 (2021)

[5] Tamura, Y., Zhang, D., Umeda, N., Sakashita, K.: Load forecasting using grey dynamic model. Journal of Grey System **4**(1) (1992)

[6] Olabode, O., Amole, O., Ajewole, T., Okakwu, I.: Medium-term load forecasting in a nigerian electricity distribution region using regression analysis techniques. In: 2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS), pp. 1–5 (2020). IEEE

[7] AlRashidi, M., El-Naggar, K.: Long term electric load forecasting based on particle swarm optimization. Applied Energy **87**(1), 320–326 (2010)

[8] Oliveira, E.M., Oliveira, F.L.C.: Forecasting mid-long term electric energy consumption through bagging arima and exponential smoothing methods. Energy **144**, 776–788 (2018)

[9] Kim, D.-H., Lee, E.-K., Qureshi, N.B.S.: Peak-load forecasting for small industries: A machine learning approach. Sustainability **12**(16), 6539 (2020)

[10] Tang, L., Wang, X., Wang, X., Shao, C., Liu, S., Tian, S.: Long-term electricity consumption forecasting based on expert prediction and fuzzy bayesian theory. Energy **167**, 1144–1154 (2019)

[11] Ju-Long, D.: Control problems of grey systems. Systems & control letters **1**(5), 288–294 (1982)

[12] Morita, H., Zhang, D.-P., Tamura, Y.: Long-term load forecasting using grey system theory. Electrical engineering in Japan **115**(2), 11–20 (1995)

[13] Kang, J., Zhao, H.: Application of improved grey model in long-term load forecasting of power engineering. Systems Engineering Procedia **3**, 85–91 (2012)

[14] Dudek, G.: Pattern-based local linear regression models for short-term load forecasting. Electric power systems research **130**, 139–147 (2016)

[15] Amjady, N.: Short-term bus load forecasting of power systems by a new hybrid method. IEEE Transactions on Power Systems **22**(1), 333–341 (2007)

[16] Shafie-Khah, M., Moghaddam, M.P., Sheikh-El-Eslami, M.: Price forecasting of day-ahead electricity markets using a hybrid forecast method. Energy Conversion and Management **52**(5), 2165–2169 (2011)

[17] Da-Hai, Z., Shi-Fang, J., Yan-Qiu, B.I., Gui-Bin, Z.: Study of power system load forecast based on wavelet neural networks. Electric Power Automation Equipment (2003)

[18] Barman, M., Choudhury, N.B.D.: Season specific approach for short-term load forecasting based on hybrid fa-svm and similarity concept. Energy **174**, 886–896 (2019)

[19] Hanmandlu, M., Chauhan, B.K.: Load forecasting using hybrid models. IEEE Transactions on Power Systems **26**(1), 20–29 (2010)

[20] Imani, M.: Electrical load-temperature cnn for residential load forecasting. Energy **227**, 120480 (2021)

[21] Zheng, J., Xu, C., Zhang, Z., Li, X.: Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network. In: 2017 51st Annual Conference on Information Sciences and Systems (CISS), pp. 1–6 (2017). IEEE

[22] Raza, M.Q., Khosravi, A.: A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renewable and Sustainable Energy Reviews **50**, 1352–1372 (2015)

[23] Mocanu, E., Nguyen, P.H., Gibescu, M., Kling, W.L.: Deep learning for estimating building energy consumption. Sustainable Energy, Grids and Networks **6**, 91–99 (2016)

[24] Liu, M., Qin, H., Cao, R., Deng, S.: Short-term load forecasting based on improved tcn and densenet. IEEE Access **10**, 115945–115957 (2022)

[25] Han, L., Peng, Y., Li, Y., Yong, B., Zhou, Q., Shu, L.: Enhanced deep networks for short-term and medium-term load forecasting. Ieee Access **7**, 4045–4055 (2018)

[26] Ahmad, T., Chen, H.: Potential of three variant machine-learning models for forecasting district level medium-term and long-term energy demand in smart grid environment. Energy **160**, 1008–1020 (2018)

[27] Wang, X., Ahn, S.-H.: Real-time prediction and anomaly detection of electrical load in a residential community. Applied Energy **259**, 114145 (2020)

[28] Nepal, B., Yamaha, M., Yokoe, A., Yamaji, T.: Electricity load forecasting using clustering and arima model for energy management in buildings. Japan Architectural Review **3**(1), 62–76 (2020)

[29] Liu, Y., Wang, W., Ghadimi, N.: Electricity load forecasting by an improved forecast engine for building level consumers. Energy **139**, 18–30 (2017)

[30] Khan, Z.A., Hussain, T., Ullah, A., Rho, S., Lee, M., Baik, S.W.: Towards efficient electricity forecasting in residential and commercial buildings: A novel hybrid cnn with a lstm-ae based framework. Sensors **20**(5), 1399 (2020)

[31] Hoori, A.O., Al Kazzaz, A., Khimani, R., Motai, Y., Aved, A.J.: Electric load forecasting model using a multicolumn deep neural networks. IEEE Transactions on Industrial Electronics **67**(8), 6473–6482 (2019)

[32] Yang, Y., Che, J., Deng, C., Li, L.: Sequential grid approach based support vector regression for short-term electric load forecasting. Applied energy **238**, 1010–1021 (2019)

[33] Ionescu, M., Păun, G., Yokomori, T.: Spiking neural p systems. Fundamenta informaticae **71**(2-3), 279–308 (2006)

[34] Liu, Q., Long, L., Yang, Q., Peng, H., Wang, J., Luo, X.: Lstm-snp: A long short-term memory model inspired from spiking neural p systems. Knowledge-Based Systems **235**, 107656 (2022)

[35] Dehalwar, V., Kalam, A., Kolhe, M.L., Zayegh, A.: Electricity load forecasting

for urban area using weather forecast information. In: 2016 IEEE International Conference on Power and Renewable Energy (ICPRE), pp. 355–359 (2016). IEEE

[36] Boroojeni, K.G., Amini, M.H., Bahrami, S., Iyengar, S., Sarwat, A.I., Karabasoglu, O.: A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon. Electric Power Systems Research **142**, 58–73 (2017)

[37] Wilms, H., Cupelli, M., Monti, A.: Combining auto-regression with exogenous variables in sequence-to-sequence recurrent neural networks for short-term load forecasting. In: 2018 IEEE 16th International Conference on Industrial Informatics (INDIN), pp. 673–679 (2018). https://doi.org/10.1109/INDIN.2018.8471953