

Featuring periodic correlations via dual granularity inputs structured RNNs ensemble load forecaster

Lyuzerui Yuan | Jinghuan Ma | Jie Gu | Honglin Wen | Zhijian Jin

Department of Electronic Information and Electrical Engineering, Research Center for Big Data Engineering and Technologies, Shanghai Jiao Tong University, Shanghai, China

Correspondence

Jie Gu, Department of Electronic Information and Electrical Engineering, Research Center for Big Data Engineering and Technologies, Shanghai Jiao Tong University, Minhang District, Shanghai 200240, China.
 Email: gujie@sjtu.edu.cn

Funding information

Key Project of Shanghai Science and Technology Committee, Grant/Award Number: 18DZ1100303; Shanghai Sailing Program, Grant/Award Number: 19YF1423700; the National Key Research and Development Program of China, Grant/Award Number: 2016YFB0900100

Peer Review

The peer review history for this article is available at <https://publons.com/publon/10.1002/2050-7038.12571>.

Summary

With the development of Energy Internet and the widespread use of smart meters, the application of user-side data has been attracting more and more attention, one of which is residential short-term load forecasting (STLF). The ability of traditional prediction methods is limited due to a large amount of fine-grained data with high uncertainty, and many researchers have applied deep learning to STLF thereby. However, perspectives to study and develop deep neural network forecasters are yet to be freed from time domain. Frequency domain analysis provides a sufficient path to assess the periodic characteristics of load. By insights from the spectrum of load data, significant periodic fluctuation characteristics of load are revealed. Based on this, a dual time granularity inputs structured RNNs ensemble STLF model is proposed, which unites two independent input-featured RNNs to learn more time-frequency characteristics implied in fine-grained load. One RNN is mainly responsible for mining time-domain features of load, the other focuses on learning frequency-domain features. Case study indicates a positive correlation between capturing more spectrum characteristics and performing a better prediction over accepted criteria, which has been achieved by the proposed model. In addition, experiments show that the proposed forecaster can improve both individual household and aggregated load STLF accuracy.

KEY WORDS

deep learning, ensemble model, recurrent neural network, short-term load forecasting, spectrum analysis, time series

1 | INTRODUCTION

Load forecasting has been an essential work to ensure the security and stability of power system. In recent years, residential short-term load forecasting (STLF) has attracted the attention of many researchers, which is important for Demand Response Management,¹ Energy Internet,² and so on. However, compared with industrial and commercial

List of symbols and abbreviations: $x_{(d,t)}$, load data at time point (d,t) ; (d,t) , a subscript for time point means the t th sampling point of the d th day; D , number of days of data; T , number of sampling points in a day; m , sequential length; n , periodic length; $\mathbf{X}_{(d,t),m,S}$, sequential input at (d,t) ; $\mathbf{X}_{(d,t),n,P}$, periodic input at (d,t) ; $\hat{y}_{(d,t),S}$, the output of S-RNN at (d,t) ; $\hat{y}_{(d,t),P}$, the output of P-RNN at (d,t) ; $\hat{y}_{(d,t),D}$, the output of DI-RNN at (d,t) ; $W_i, W_h^1, W_h^2, W^{12}$, the weight parameters of RNN with two hidden layers; b^1, b^2 , the bias parameters of RNN with two hidden layers; W_{fc1}, W_{fc2} , the weight parameters of one fully connection layer after RNN; b_{fc1}, b_{fc2} , the bias parameters of one fully connection layer after RNN; W_{bp1}, W_{bp2} , the weight parameters of BPNN in DI-RNN; b_{bp1}, b_{bp2} , the bias parameters of BPNN in DI-RNN.

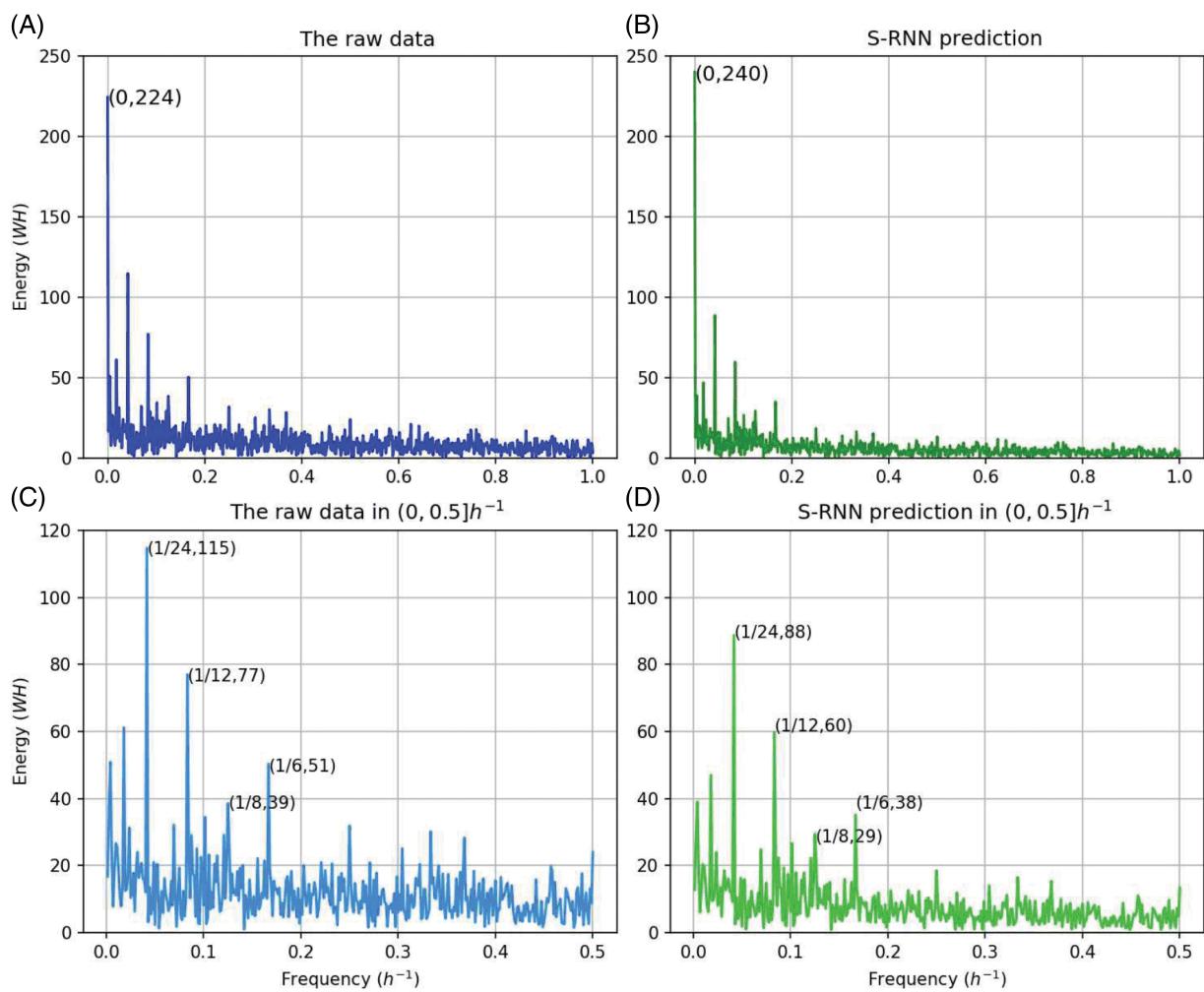


FIGURE 1 Spectrums of user 1's one-month electricity consumption data sampled every 30 minutes and corresponding S-RNN prediction. It can be seen from subfigure, A, that the user's spectrum does not show obvious periodicity when the frequency is greater than $0.5/h^{-1}$. For convenience of observation, we intercept the spectrum of subfigures, A and B, with a frequency range $(0, 0.5]/h^{-1}$ to obtain subfigures, C and D, respectively

load, residential load has strong uncertainty and volatility to make accurate residential STLF difficult.³ With the popularization of smart meters,⁴ a large amount of residential load data becomes available to mine the meter-level load characteristics, since it contains extensive information on residential power consumption behavior.

To our best knowledge, the traditional ideas about unearthing residential load characteristics can be divided into four categories: clustering,⁵⁻⁸ spectral analysis,^{9,10} feature selection^{11,12} and adding external information.^{13,14} (a) Clustering: Jungsuk Kwac et al proposed a load shape encoding system to cluster various loads accurately⁸; Alzate and Sinn took kernel clustering to obtain different residential clusters and then established a prediction model for each cluster.¹⁵ (b) Spectral analysis: Ghofrani et al used spectral analysis to get the load shaping filter and then utilized Kalman filtering for STLF⁹; wavelet decomposition was used to mine frequency characteristic of load by Chen et al.¹⁶ (c) Feature selection: Pirbazari et al applied four feature selection methods for SLTF, that is, F-regression, Mutual Information, Recursive Feature Elimination and Elastic Net¹⁷; based on extreme learning machine, the conditional mutual information was developed to select a compact set of input variables by Song Li et al for increased forecasting performance.¹⁸ (d) Adding external information: a multiple regression prediction model was proposed by Pu Wang et al, using many kinds of external information including temperature, week, and so on¹⁹; Cao et al considered meteorological conditions in their proposed STLF model to improve the prediction accuracy.²⁰ Although these methods can improve the prediction accuracy, clustering and spectral analysis have limited efficiency to massive data; the process of feature selection is usually time-consuming; external variables lead to a large increase in the amount of data, and the work by

FIGURE 2 An example of dual granularity inputs

Date \ Time Consumption (WH)	12:00	12:30	13:00	13:30	14:00	14:30
1-Apr	76	120	136	74	154	94
2-Apr	288	340	230	312	242	300
3-Apr	190	140	84	74	148	112
4-Apr	236	158	76	150	104	112
5-Apr	254	310	222	290	222	246
6-Apr	268	410	436	252	138	92
7-Apr	194	74	132	162	76	104

Sequential Input

Periodic Input

Expected Output

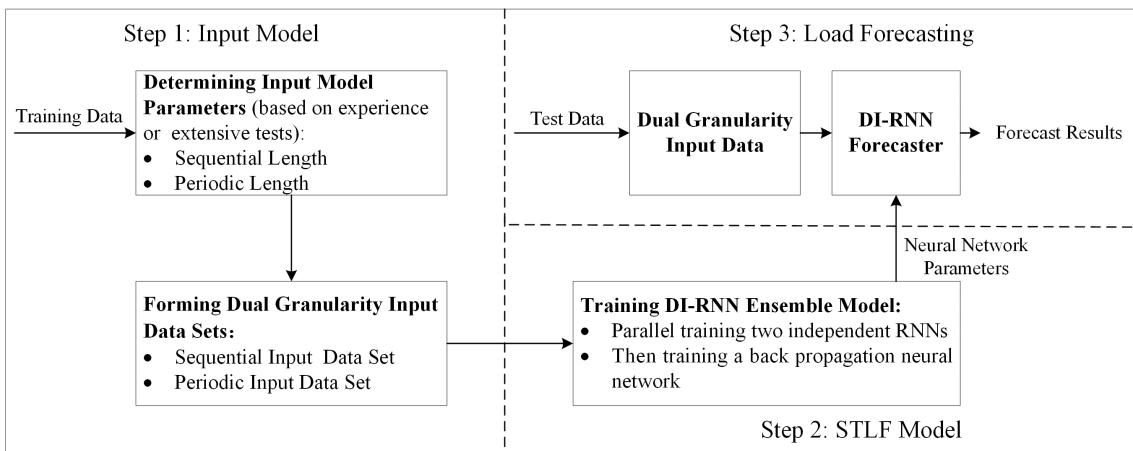


FIGURE 3 Residential STLF process based on DI-RNN ensemble model

James W. Taylor pointed out that weather variables have little effect on short prediction intervals such as 60-minute interval or less.²¹

The emergence of Deep Learning provides researchers a new idea for STLF, which can establish brain-like neural networks to learn the characteristics of massive data. Among the mainstream Deep Learning networks, Recurrent Neural Network (RNN)²² with the ability to memorize short-term related information of time series, has drawn much attention in STLF.^{23,24} Ljubisa Sehovac and Katarina Golinger studied sequence to sequence RNN prediction model combined with attention mechanism²⁵; Heng Shi et al proposed a pooling RNN forecaster that batched historical load data of neighbors into a pool of RNN inputs.²⁶ There are also other studies that combine traditional methods with Deep Learning. For instance, Mingyang Sun et al. applied clustering and Bayesian RNN for residential STLF.⁵

The existing RNN based STLF models only utilize the time-domain characteristics of residential load. However, the characteristics of load are not only manifesting in time domain but also in frequency domain. This means we can further improve the capability of neural network forecasters by spectrum analysis, in addition to time-domain analysis. There is an intuitive example. In Figure 1, the sub-figures (A and C) present the spectrum of a residential electricity consumer's monthly load profile,²⁷ of which the sampling period is 0.5 hour; the right ones (B and C) present the spectrum of a conventional RNN STLF result based on the same consumer data. The input structure of the RNN that contains the time-domain characteristics of load, that is, m consecutive points received by the RNN to predict the $(m + 1)$ th point, is named sequential input hereafter. The RNN is named sequential RNN (S-RNN).

We can infer significant knowledge in terms of behavioral correlation from the spectrum. For spectrum of the raw data, energy is remarkable at $\frac{1}{24} h^{-1}$, which means consumer behavior contains a high ratio of 24-hour cycle component. That is, two load points with a 24-hour interval are significantly correlated. Besides, 12h, 6h and so on are less strong

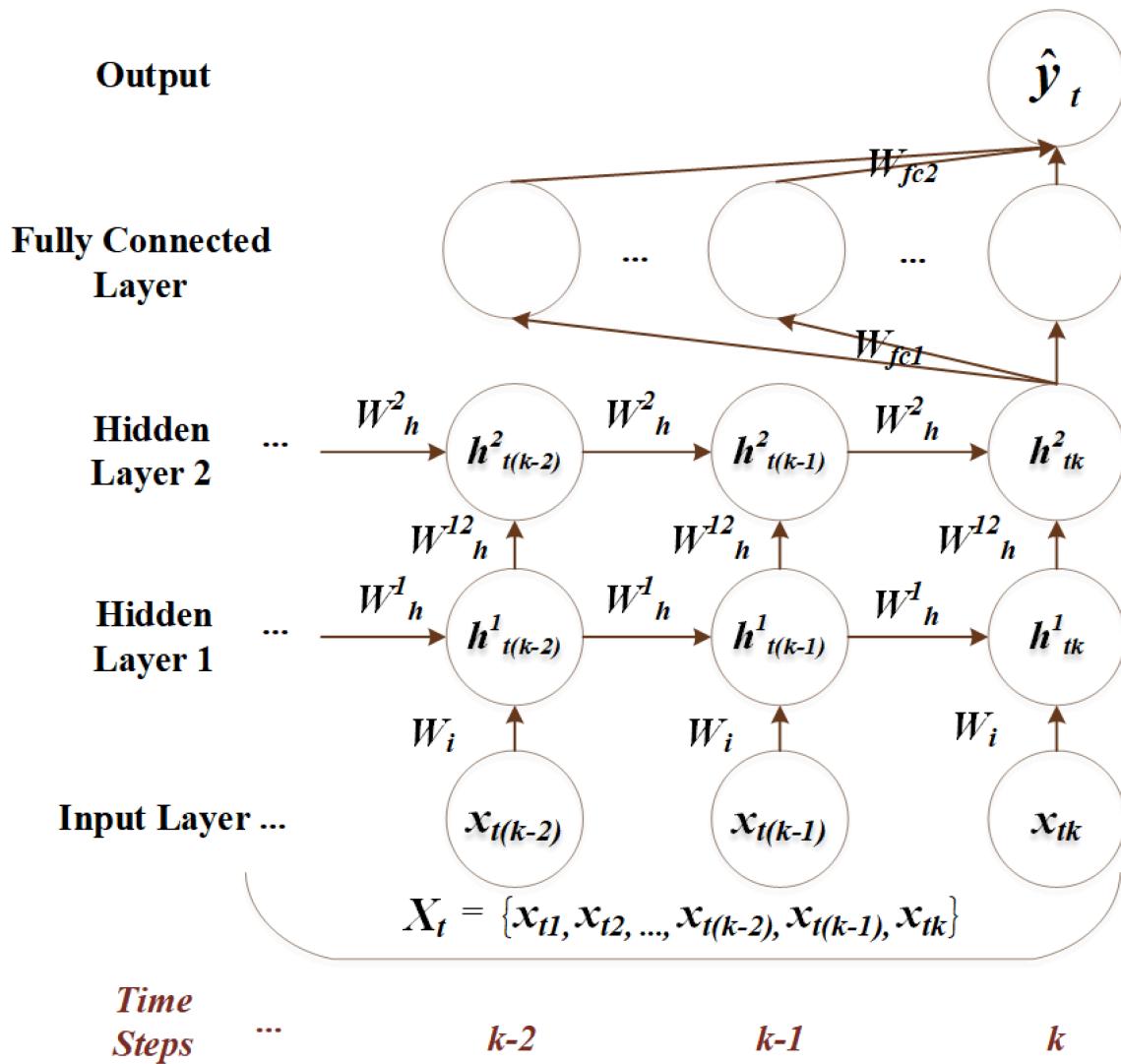


FIGURE 4 Structure of a fully connected RNN with two hidden layers

yet relatively significant cycles. Comparing the two spectrums, the energies of the prediction are less than those of the raw data at those significant frequencies, while the prediction has more constant component according to Figure 1A,B. It is inferred that conventional RNN has limited ability to extract periodic information, or from the perspective of feature engineering, the sequential input does not help the neural networks to strengthen the periodic correlation between certain points.

For a more accurate STL, this paper first proposes to strengthen the ability of conventional RNN by recording significant periodic correlations implied in the load data. On the basis of the analysis above, we design a new input structure, that is, dual granularity inputs (DI), based on which to develop an RNNs ensemble STL model by two RNN models focusing on unearthing the time-domain and frequency-domain characteristics of load respectively. As a preliminary attempt, only the 24-hour cycle is selected as a characteristic period with regard to the design of input, as the energy of the raw data is most remarkable at $\frac{1}{24}h^{-1}$ in Figure 1C. According to the periodic fluctuation characteristics of load, a periodic input reflecting the frequency characteristics of data is proposed, which uses the closest previous n (periodic length) load points with a 24-hour interval from the target point. Then periodic input and sequential input are together to form DI structure. Figure 2 provides an example of DI whose sequential length is 4 and periodic length is 5. Corresponding to DI, an ensemble model, that is, DI structured RNNs (DI-RNN) is constructed that combines two independent-trained input-featured neural networks: S-RNN and periodic input structured RNN (P-RNN).

Experiments based on real single and aggregated residential load data from Australia²⁷ show that DI-RNN has an excellent prediction performance over two accepted criteria, that is, the overall error and the error distribution, with

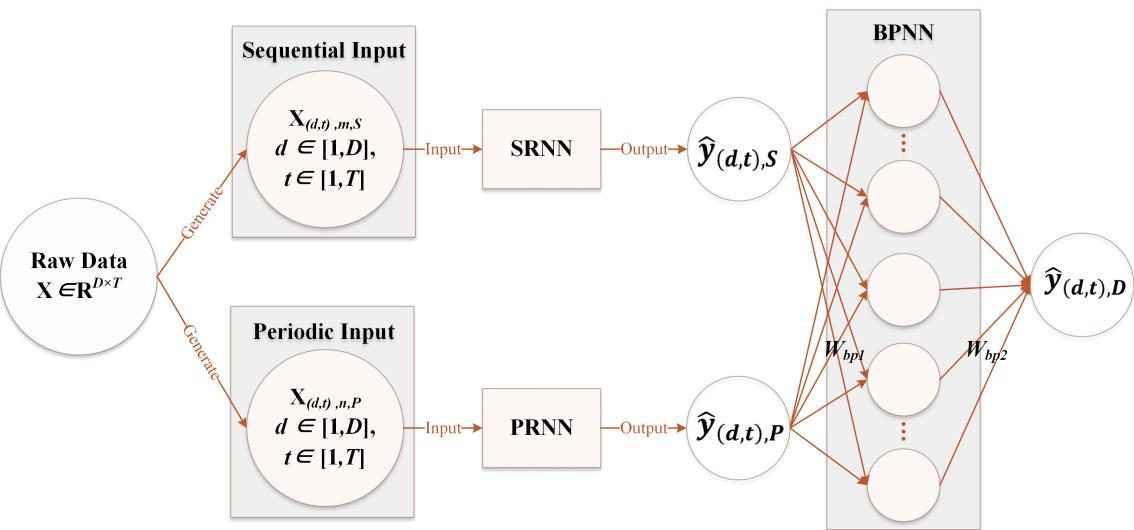


FIGURE 5 Structure of DI-RNN ensemble forecaster

TABLE 1 The parameters setting of five benchmarks

Benchmarks	Parameters
SVR	Input length = 96; kernel = "rbf"; C = 1 000 000; gamma = 0.005
GBDT	Input length = 96; learning rate = 0.1; estimator number = 150; max depth = 4
BPNN	Input length = 192; hidden layers = {first layer: 24 neurons; second layer: 12 neurons; third layer = 12 neurons}; learning rate = 0.005; optimizer = Adam
RF	Input length = 96; estimator number = 100; max depth = 3
Wavelet +ARIMA	Wavelet = Daubechies; level = 2; information criterion = AIC

limited extra required spatial complexity. Moreover, a new metric named spectrum similarity (SS) is introduced to evaluate the performance of the STLF model in frequency domain, based on which the learning ability for periodic correlations of the proposed model is confirmed. Hence, the proposed DI-RNN load forecaster is verified to succeed in learning more periodic correlations of the load data to achieve improved performance, according to common criteria of prediction accuracy. The main contributions of this study are as follows:

1. We set analysis of fine-grained residential load free from time domain. Through the spectrum analysis, the high uncertainty residential electricity consumption data reveals its regular periodic fluctuation.
2. A succinct RNN based input structure is proposed, which takes into account the time-domain and frequency-domain characteristics of load.
3. A new ensemble forecaster is designed that combines two independent trained input-featured RNNs to mine residential load characteristics from time and frequency domains. Case study shows that the proposed model can improve prediction accuracy not only for individual household but also for aggregated load.

The remainder of the paper is organized as follows. In Section 2, DI and DI-RNN are introduced in detail. Section 3 presents case study to evaluate the proposed forecaster and display its improved performance. Conclusion is drawn in Section 4.

2 | DUAL GRANULARITY INPUTS STRUCTURED RNNs

The residential STLF process based on the proposed model is shown in Figure 3, in which step 1 and 2 will be mainly introduced. In this section, the input model, that is, dual granularity inputs (DI), is first developed. A fully connected RNN is then defined. The proposed DI-RNN ensemble model and three evaluation criteria are finally presented.

TABLE 2 The final prediction results of three users from 1 December 2013 to 30 March 2014

Architecture	User 1				User 2				User 3				User 4				User 5			
	MAE (WH)	EAIP	SS	MAE (WH)	EAIP	SS	MAE (WH)	EAIP	SS	MAE (WH)	EAIP	SS	MAE (WH)	EAIP	SS	MAE (WH)	EAIP	SS		
S-RNN/96	96.94	0.43	140.63	76.46	0.37	80.28	138.50	0.52	82.31	41.48	0.57	38.28	60.32	0.22	52.71					
DI-RNN/(96,2)	90.58	0.52	109.26	66.81	0.46	63.12	131.94	0.58	67.59	33.93	0.70	20.34	50.66	0.40	21.11					
S-RNN/192	95.14	0.44	133.38	74.42	0.41	77.11	142.51	0.44	115.98	43.41	0.57	45.17	58.48	0.22	46.84					
DI-RNN/(192,4)	88.07	0.54	106.61	62.64	0.49	61.94	134.54	0.53	74.34	36.56	0.65	24.79	49.72	0.41	22.11					
S-RNN/288	97.70	0.42	132.47	74.29	0.42	72.62	148.13	0.40	113.47	45.23	0.53	47.28	58.80	0.22	44.29					
DI-RNN/(288,6)	88.67	0.54	114.17	61.51	0.48	57.32	141.23	0.49	72.82	38.12	0.63	26.77	47.25	0.43	20.87					
S-RNN/384	102.26	0.38	149.74	72.36	0.42	72.70	166.77	0.33	132.74	43.52	0.55	45.52	57.32	0.23	43.62					
DI-RNN/(384,8)	92.67	0.49	120.47	61.27	0.50	59.42	152.68	0.41	119.74	38.24	0.63	26.51	48.22	0.41	21.90					
Average improvement from S-RNN to DI-RNN(%)	8.15	24.60	18.93	15.24	19.06	20.09	5.86	15.76	24.85	15.45	18.07	44.28	16.63	87.16	53.85					
SVR	93.42	0.53	143.14	63.14	0.44	78.75	141.68	0.47	106.90	35.88	0.64	29.6	49.1	0.43	52.88					
Average improvement from SVR to DI-RNN(%)	3.66	1.23	21.32	0.12	10.05	23.23	1.12	6.92	21.78	2.33	2.24	16.88	0.28	4.07	59.35					
GBDT	93.78	0.49	164.06	63.50	0.44	78.83	142.80	0.43	117.19	40.78	0.60	37.32	51.10	0.42	55.60					
Average improvement from GBDT to DI-RNN(%)	4.03	7.07	31.35	0.69	10.05	23.31	1.89	16.86	28.64	9.97	9.05	34.07	4.18	1.79	61.34					
BPNN	90.22	0.52	149.23	64.64	0.43	85.69	144.42	0.416	124.70	39.53	0.611	31.88	48.82	0.44	53.64					
Average improvement from BPNN to DI-RNN(%)	0.25	0.29	24.53	2.45	11.83	29.45	2.99	20.79	32.94	7.12	7.06	22.83	0.29	6.25	59.92					
RF	106.08	0.46	147.26	69.14	0.42	80.44	152.00	0.42	123.67	49.34	0.50	45.73	52.45	0.42	54.70					
Average improvement from RF to DI-RNN(%)	15.16	13.59	23.52	8.79	15.29	24.84	7.83	19.64	32.38	25.59	30.87	46.20	6.65	1.79	60.70					
Wavelet+ARIMA	111.62	0.33	173.22	83.41	0.34	96	179.30	0.28	146.85	48.7	0.47	41.23	65.26	0.20	54.72					
Average improvement from Wavelet +ARIMA to DI-RNN(%)	19.50	57.75	34.90	24.39	42.42	37.02	21.73	86.60	43.11	24.61	39.22	40.33	24.67	106.25	60.70					

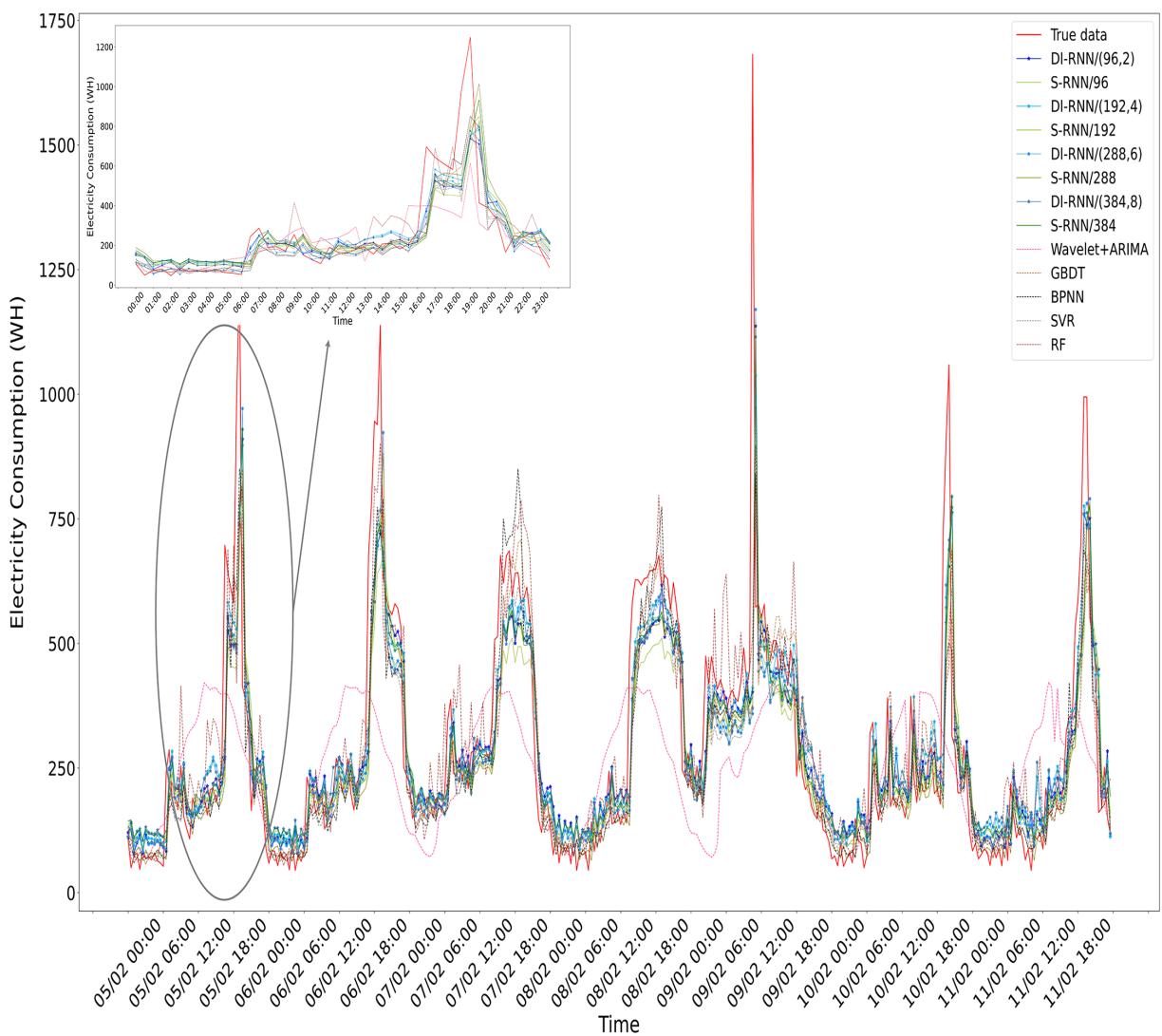


FIGURE 6 The electricity consumption predictions of user 1 from 5 February 2014 to 11 February 2014, based on different STLF models. We enlarge the daily forecast profiles on 5 February 2014 for a clearer comparison of the results

2.1 | Dual granularity inputs

Let $\mathbf{X} = \{x_{(d,t)} \mid d \in [1, D], t \in [1, T]\}$ denotes a time series of load where D represents the number of intended days and T is the number of sampling points in a day. We can locate a load in accordance with the subscript (d, t) which refers to the t th sampling point of the d th day. The corresponding calculation rule is defined as follows:

1. $(d - n, t)$ refers to the t th time point on the $(d - n)$ th ($d > n$) day.
2. $(d, t - s)$ refers the $(t - s)$ th ($t > s$) time point on the d th day.
3. $(d, t) - i$ refers to the i th former time point of (d, t) .

Dual granularity inputs (DI) consists of two input structures: sequential input and periodic input. To predict $x_{(d,t)}$, sequential input, which focuses on the load near the target time point, adopts m (sequential length) consecutive points before $x_{(d,t)}$. The corresponding input is expressed as:

$$\mathbf{X}_{(d,t),m,S} = \{x_{(d,t)-m}, \dots, x_{(d,t)-1}\}. \quad (1)$$

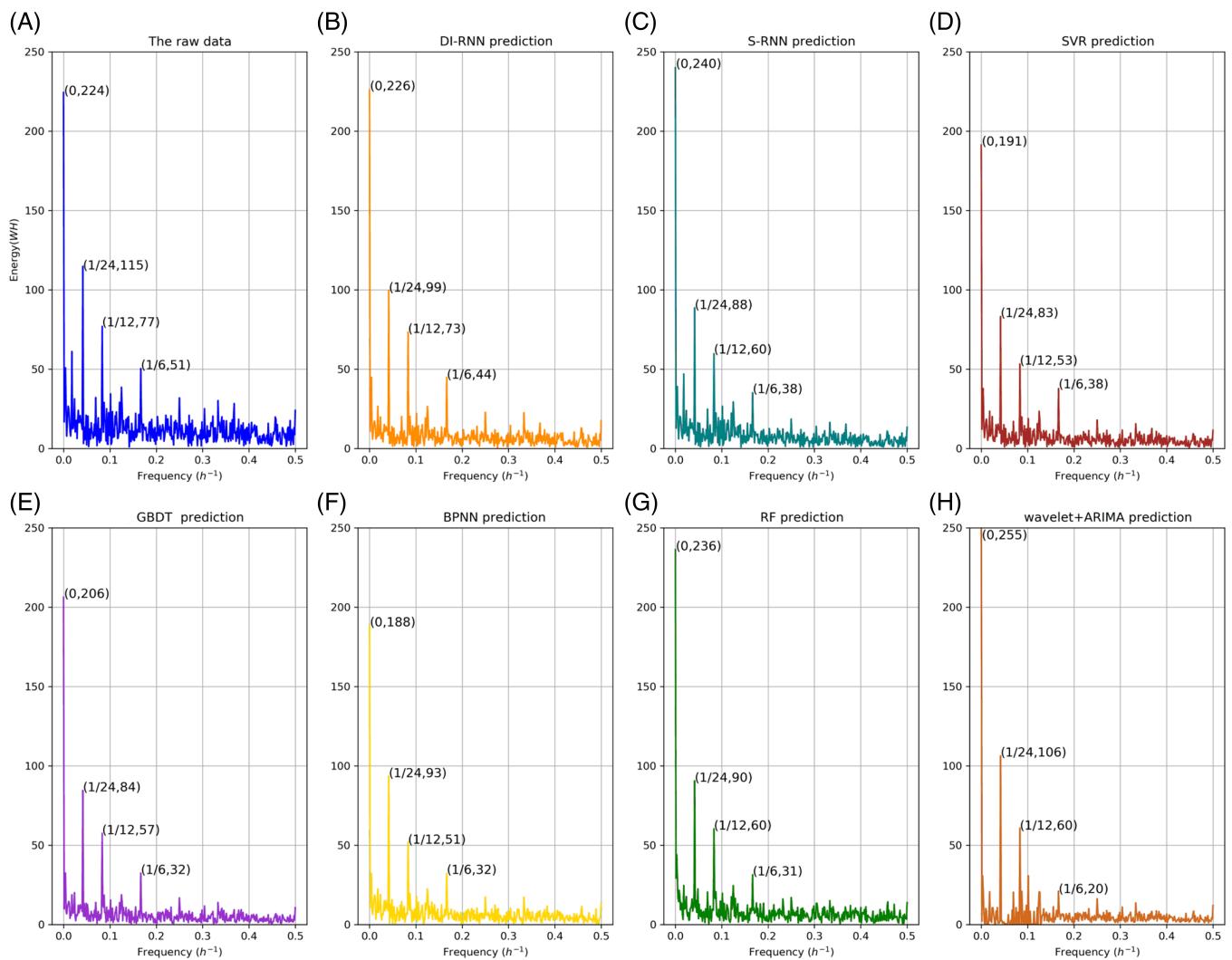


FIGURE 7 Spectral comparison of user 1' raw data, DI-RNN prediction, S-RNN prediction, SVR prediction, GBDT prediction, BPNN prediction, RF prediction and “Wavelet+ARIMA” prediction

From Figure 1, we can see that sequential input is limited to improve the ability of prediction models to record significant periodic correlations. Periodic input is proposed to strengthen the extraction of periodic information, that is, resamples n (periodic length) previous points of $x_{(d,t)}$ with a 24-hour cycle. The periodic input for $x_{(d,t)}$ is denoted as:

$$\mathbf{X}_{(d,t),n,P} = \{x_{(d-n,t)}, \dots, x_{(d-1,t)}\}. \quad (2)$$

Thus, DI is determined by a pair of parameters (m, n).

2.2 | A fully connected recurrent neural network

Recurrent neural network (RNN) contains an input layer, hidden layers and an output layer. The input layer refers to the input data. The output layer is responsible for generating the final results. Hidden layers do not accept external data directly or generate the final results, but learn and extract data characteristics. RNN establishes weighted connections among neurons of the same layer on hidden layers, which connects association information over a period of time. Figure 4 shows the structure of a fully connected RNN with two hidden layers (FC RNN). We define an FC RNN as:

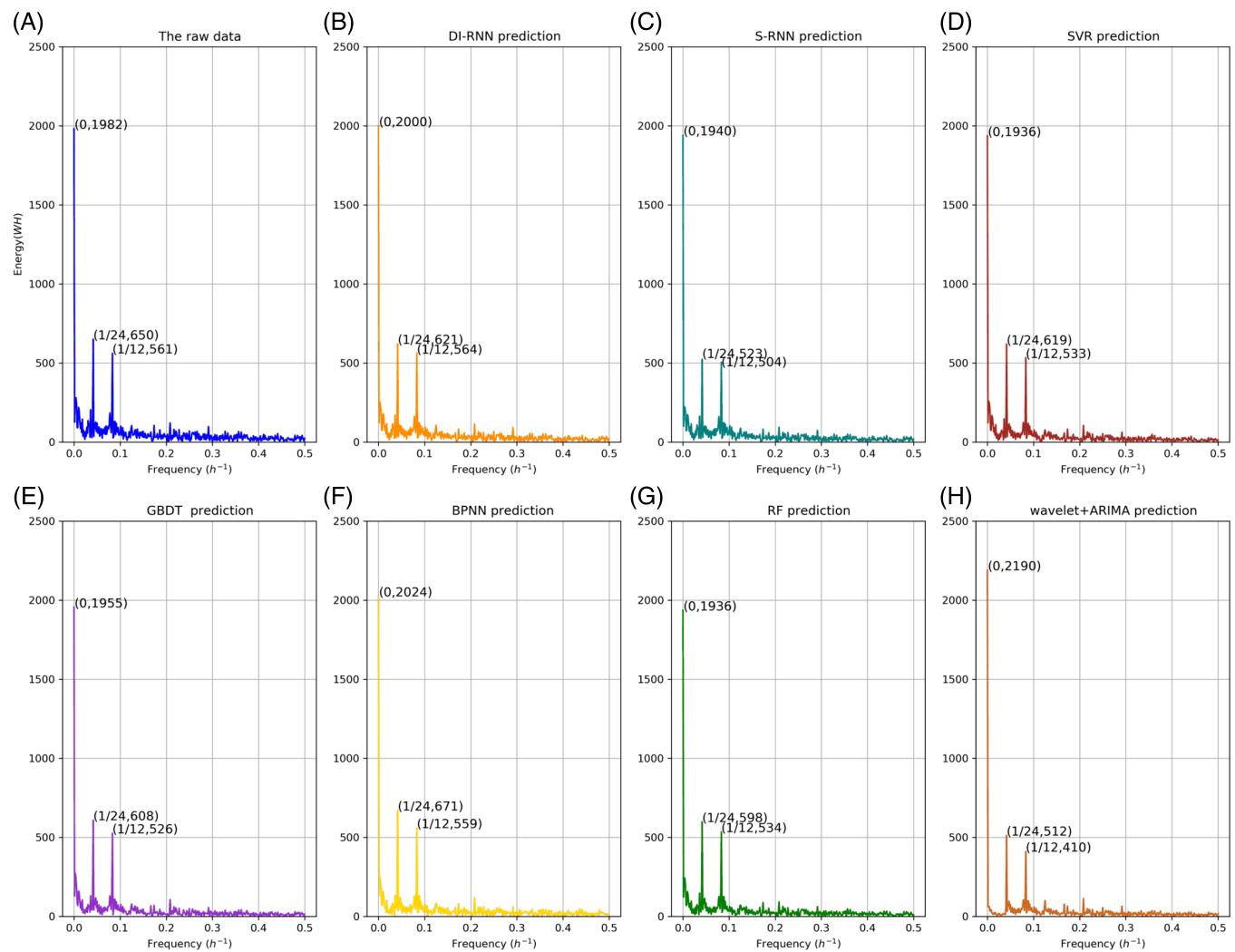


FIGURE 8 Spectral comparison of the aggregated data, DI-RNN prediction, S-RNN prediction, SVR prediction, GBDT prediction, BPNN prediction, RF prediction and “Wavelet+ARIMA” prediction

$$\hat{y} = \text{FCRNN}^G(\mathbf{X}), \quad (3)$$

where $\mathbf{X} = \{x_1, x_2, \dots, x_k\}$ represents the input which is a time series of length k ; k determines the number of time steps of RNN; \hat{y} represents the output; G denotes a set of featured parameters. At time t , the output of the first hidden layer is:

$$h_{tk}^1 = \tanh\left(W_i x_{tk} + W_h^1 h_{t(k-1)}^1 + b^1\right), \quad (4)$$

and the output of the second hidden layer is:

$$h_{tk}^2 = \tanh\left(W^{12} h_{tk}^1 + W_h^2 h_{t(k-1)}^2 + b^2\right), \quad (5)$$

where W_i , W_h^1 , W^{12} and W_h^2 are the weight parameters of RNN, b^1 and b^2 are the bias parameters. The output is acquired after passing through a fully connection layer, which is expressed as:

TABLE 3 The final prediction results of aggregated load from 1 December 2013 to 30 March 2014

Architecture	Aggregated load		
	MAE (WH)	EAIP	SS
S-RNN/96	412.00	0.70	321.15
DI-RNN/(96,2)	394.25	0.76	215.51
S-RNN/192	407.55	0.70	324.80
DI-RNN/(192,4)	387.10	0.77	212.72
S-RNN/288	405.35	0.72	318.51
DI-RNN/(288,6)	398.87	0.75	217.20
S-RNN/384	435.37	0.65	320.85
DI-RNN/(384,8)	386.60	0.77	208.57
Average improvement from S-RNN to DI-RNN(%)	5.53	10.30	33.55
SVR	392.30	0.75	360.08
Average improvement from SVR to DI-RNN(%)	1.33	3.04	43.67
GBDT	401.50	0.71	375.36
Average improvement from GBDT to DI-RNN(%)	2.32	7.40	43.07
BPNN	397.00	0.74	379.24
Average improvement from BPNN to DI-RNN(%)	1.33	3.04	43.67
RF	407.20	0.71	384.50
Average improvement from RF to DI-RNN(%)	3.81	7.39	44.40
Wavelet+ARIMA	552.48	0.56	426.41
Average improvement from Wavelet+ARIMA to DI-RNN(%)	29.17	45.10	49.90

$$\hat{y}_t = \text{FCRNN}^G(X_t) = \text{ReLU}(W_{fc2}\text{ReLU}(W_{fc1}h_{tk}^2 + b_{fc1}) + b_{fc2}). \quad (6)$$

Here, $G = \{W_i, W_h^1, W_h^2, W_h^{12}, b^1, b^2, W_{fc1}, W_{fc2}, b_{fc1}, b_{fc2}\}$ and ReLU is an activation function defined as:

$$\text{ReLU}(x) = \max(0, x), \quad (7)$$

where if $x \leq 0$, the derivative of $\text{ReLU}(x)$ is 0; if $x > 0$, the derivative of $\text{ReLU}(x)$ is 1.

2.3 | Dual granularity inputs structured RNNs ensemble model

Dual granularity inputs structured RNNs ensemble model (DI-RNN) takes RNN as basic model for prediction. There are three parts in DI-RNN: sequential input structured RNN model (S-RNN), periodic input structured RNN model (P-RNN) and back propagation neural network (BPNN).²⁸ The inputs and training processes of S-RNN and P-RNN are independent, which means that parallel computing can be performed to greatly reduce runtime. BPNN fixes the parameters of the two and carries out its own training as shown in Figure 5.

For convenience, let SRNN and PRNN represent FC RNN networks with sequential input and periodic input, respectively. The output of S-RNN at (d, t) can be expressed as:

$$\hat{y}_{(d,t),S} = \text{SRNN}^{G_S}(X_{(d,t),m,S}), \quad (8)$$

where G_S is the parameter set of SRNN and $X_{(d,t),m,S}$ is defined in (1). Similarly, the output at (d, t) of P-RNN is denoted as:

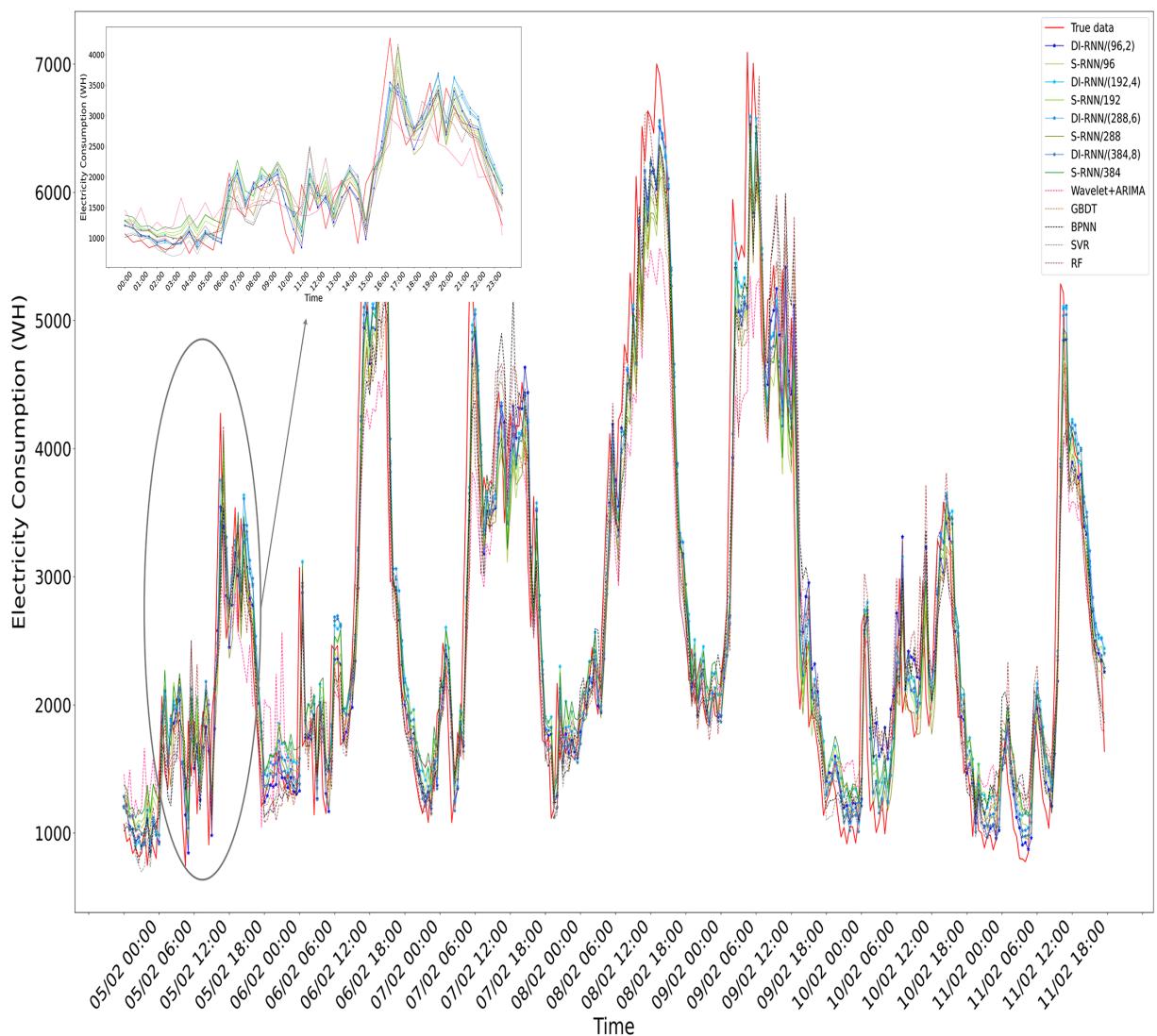


FIGURE 9 The electricity consumption predictions of aggregated users from 5 February 2014 to 11 February 2014, based on different STLF models. The forecast profiles on 5 February 2014 are enlarged for a clearer comparison

$$\hat{y}_{(d,t),P} = \text{PRNN}^{G_P}(X_{(d,t),n,P}), \quad (9)$$

where G_P is the parameter set of PRNN and $X_{(d,t),n,P}$ is defined in (2).

As an independent neural network, BPNN is used to synthesize the results of S-RNN and P-RNN. It takes the outputs at (d, t) of S-RNN and P-RNN $\{\hat{y}_{(d,t),S}, \hat{y}_{(d,t),P}\}$ as the input to generate the final result which is calculated as:

$$\hat{y}_{(d,t),D} = \text{ReLU}\left(W_{bp2}\text{ReLU}\left(W_{bp1}\left[\hat{y}_{(d,t),S}, \hat{y}_{(d,t),P}\right] + b_{bp1}\right) + b_{bp2}\right), \quad (10)$$

where W_{bp1} , W_{bp2} , b_{bp1} , b_{bp2} are parameters of BPNN.

In every training, S-RNN, P-RNN and BPNN are trained independently, which can be further indicated by the loss functions described below.

S-RNN uses back-propagation through time (BPTT) algorithm²² to converge according to the loss function which is expressed as:

$$L_s = \sqrt{\frac{1}{D \times T} \sum_{d=1}^D \sum_{t=1}^T \left(x_{(d,t)} - \hat{y}_{(d,t),S} \right)^2}, \quad (11)$$

where $\hat{y}_{(d,t),S}$ is defined in (8). P-RNN also applies BPTT and denotes its loss function as:

$$L_P = \sqrt{\frac{1}{D \times T} \sum_{d=1}^D \sum_{t=1}^T \left(x_{(d,t)} - \hat{y}_{(d,t),P} \right)^2}, \quad (12)$$

where $\hat{y}_{(d,t),P}$ is given in (9). BPNN is trained independently based on the outputs of the S-RNN and the P-RNN. Back propagation (BP) algorithm²⁸ is adopted with the loss function which is calculated as:

$$L_B = \sqrt{\frac{1}{D \times T} \sum_{d=1}^D \sum_{t=1}^T \left(x_{(d,t)} - \hat{y}_{(d,t),D} \right)^2}, \quad (13)$$

where $\hat{y}_{(d,t),D}$ is defined in (10).

2.4 | Evaluation criterion

In this work, mean absolute error (MAE), error acceptance interval probability (EAIP) and spectrum similarity (SS) are applied to assess models. MAE measures the overall error of prediction calculated as:

$$\text{MAE} = \frac{1}{D \times T} \sum_{d=1}^D \sum_{t=1}^T \left| x_{(d,t)} - \hat{y}_{(d,t),D} \right|. \quad (14)$$

From the perspective of prediction errors probability distribution, we refer to the idea of confidence interval²⁹ to define an EAIP as:

$$\text{EAIP}(a, b) = P\left(a \leq \frac{\hat{y}_{(d,t),D}}{x_{(d,t)}} < b\right), \quad (15)$$

where $[a, b]$ is the error acceptance interval; EAIP(a, b) represents the probability that the ratios of predictive values and real values fall into $[a, b]$. In all experiments of Section 3, EAIP(0.7, 1.3) is abbreviated as “EAIP.”

Spectrum similarity is introduced to measure the similarity between real and predictive values in frequency domain, which can be calculated as:

$$\text{SS} = \sum_{i=1}^l (A_{real}(i) - A_{pre}(i))^2, \quad (16)$$

where $A_{real}(i)$ and $A_{pre}(i)$ represent the amplitudes of real and predictive values at corresponding signal frequency f_i , respectively. According to the definition, a smaller value indicates a higher spectrum similarity.

3 | CASE STUDY

In this section, the proposed DI-RNN is tested over real data, using a popular deep learning open source software library named TensorFlow. All experiments are based on Python and run on a computer with Intel-i7-8565U CPU.

3.1 | Dataset and setting

3.1.1 | Dataset

An open-source dataset is chosen in the experiments which is the electricity consumption benchmarks from the Australian Governments Department of Industry, Innovation and Science.²⁷ The dataset collects 25 Victorian

households' electricity consumption data with 30-minute resolution from 1 April 2012 to 31 March 2014 ($D = 730$, $T = 48$). There are three user types according to the official data description: general, controlled and generation. Five households are selected with least missing data from general type as cases, whose IDs are 2447, 3494, 4192, 5646, 9918 and numbered them user 1 to user 5, respectively. We apply the linear interpolation to fill the missing value; the three-sigma rule is used to deal with the outlier data.

In order to test the effectiveness of DI-RNN for aggregated residential load, we sum the load profiles of 10 general users to form an aggregated data set. For each household and the aggregated load, we split the data into training set and test set compliance at the ratio of 8:2. Maximum and minimum normalization is used to process all data for fast convergence of networks and improvement of model accuracy.

3.1.2 | Parameters setting

For comparison, DI-RNN and S-RNN separately apply RNN with two hidden layers of 12 neurons per layer as the basic model and the BPNN of DI-RNN has one hidden layer with five neurons per layer. The popular neural network optimizers are: Stochastic Gradient Descent, Momentum, AdaGrad, RMSProp and Adam. Adam³⁰ contains the advantages of Momentum and AdaGrad and experiments show that Adam can converge quickly. Therefore, Adam is adopted as the optimizer of all neural networks in this paper. Dropout³¹ is a good mechanism to avoid over-fitting, which refers to that each neuron is temporarily discarded from the network according to a certain probability. Through several tests, the probability of dropout is set to 0.2. The learning rates of RNN and BPNN are 0.005 and 0.008 respectively, which are selected by trial-and-error.

We focus on the prediction ability of DI-RNN and S-RNN under the same time range of the inputs, that is, the time interval between the first and last inputs of sequential input is the same as that of periodic input, which can be expressed as $m = n \times 48$ (There are 48 samples a day). To test the sensitivity of DI-RNN to input parameters, we set four sets of input parameters $(m, n) \in \{(96, 2), (192, 4), (288, 6), (384, 8)\}$.

The prediction ability of P-RNN has not attracted much attention. There are two main reasons: (a) The input of P-RNN is filtered periodically to lose a lot of adjacent information of the target load, that is, the periodic input could at most contain the lagged historical load 1 day before, while lagged half-hour historical load could be fed to S-RNN/DI-RNN. This makes the STL capability of P-RNN unsatisfactory. (b) The purpose of designing P-RNN is to strengthen the learning ability of the conventional RNN to the load periodic characteristics, rather than to predict alone. Therefore, the results of P-RNN are not shown additionally.

3.1.3 | Benchmarks setting

To demonstrate the effectiveness of DI-RNN, several state-of-the-art models are selected as benchmarks, that is, support vector regression (SVR),³² gradient boosting decision tree (GBDT),³³ BPNN and Random Forest (RF).³² Through many trials, the parameters of benchmarks are selected according to the best results, which are shown in Table 1.

The STL model combining wavelet analysis and ARIMA proposed by Cheng-Ming Lee et al.,³⁴ which is also based on the time-frequency analysis of load, is set as another benchmark. We abbreviate this benchmark as "Wavelet+ARIMA."

3.2 | Results and analyses

The results for each user are the average of 10 repeated experiments under each case illustrated in Table 2, in which DI-RNN/ (m, n) refers to DI-RNN model with the sequential length m and periodic length n and S-RNN/ m means S-RNN model with sequential length m .

3.2.1 | Time-domain evaluation of DI-RNN prediction

From Table 2, though "Wavelet+ARIMA" mines load characteristics from time-frequency perspective, the prediction results are the worst. The MAEs and EAIPs of GBDT and RF are inferior to those of DI-RNN, while the two

time-domain metrics of SVR and BPNN are close but slightly inferior to those of DI-RNN. DI-RNN is superior to S-RNN in these two measurements that MAE and EAIP are increased by 12.27%, 33.51% compared on the average of prediction results of five users respectively, that is, DI-RNN load forecaster generally has better prediction performance. Figure 6 shows the predictions of different models for user 1's electricity consumption from 5 February 2014 to 11 February 2014, as we can see, the results of DI-RNN fit the real data best.

3.2.2 | Spectrum of DI-RNN prediction

The average SS of DI-RNN is 32.40%, 28.51%, 35.74%, 33.93%, 37.53% and 43.22% better than that of S-RNN, SVR, GBDT, BPNN, RF and "Wavelet+ARIMA," respectively. It is noteworthy that even though the performance of BPNN and SVR in time domain is similar to that of DI-RNN, DI-RNN in frequency domain is quite superior to BPNN and SVR, that is, the proposed DI-RNN manages to learn more periodic correlations of load.

Taking all benchmarks, DI-RNN/(192,4) and the corresponding S-RNN/192 as an example, we perform FFT on user 1's raw data and prediction results respectively and obtain their spectra as shown in Figure 7. Obviously, the spectral ratios of DI-RNN prediction at $\frac{1}{24}h^{-1}$, $\frac{1}{12}h^{-1}$, and $\frac{1}{6}h^{-1}$ (It can be seen from the figure that these frequency points are the most significant ones in reality.) are the closest to those of raw data in all comparison models except for "Wavelet +ARIMA." Although the energy of "Wavelet+ARIMA" spectrum at $\frac{1}{24}h^{-1}$ is closest to the true situation than those of other seven model spectrums, it differs most from the real value at other obvious frequencies such as $\frac{1}{12}h^{-1}$ and $\frac{1}{6}h^{-1}$, which means "Wavelet+ARIMA" can only learn the local periodic characteristics of load. In addition, the constant component of DI-RNN prediction spectrum is also closest to the true situation, while the constant components of S-RNN, RF and "Wavelet+ARIMA" are significantly higher and those of SVR, GBDT and BPNN are lower.

Hence, we can confirm that DI-RNN is more capable of capturing periodic fluctuation information to improve prediction accuracy, whether according to the overall spectrum similarity or the significant energy points.

3.2.3 | Sensitivity analysis DI-RNN

After several tests, it is found that the results are similar when the learning rates of RNN and BPNN in DI-RNN are 0.001 to 0.008 and 0.005 to 0.008, respectively. Therefore, we will focus on the influence of input structure parameters on DI-RNN.

Although the two parameters of DI-RNN m and n determine the input of S-RNN and P-RNN respectively, we set the constraint in experiments for comparison, that is, $m = n \times 48$, which means only the influence of one parameter such as m on DI-RNN needs to be discussed. Table 2 shows that the prediction results of DI-RNN when m is 96 or 192 are better than those when m is 288 or 384, that is, the increase of m does not always bring positive gain to the predictions. Due to the insufficient learning ability of RNN for long-term information,²² the inferior predictions of S-RNN/288 and S-RNN/384 affect the results of DI-RNN/(288,6) and DI-RNN/(384,8), respectively. However, the results of the ensemble forecaster do not change greatly with the increase of m as S-RNN does, which because the results of S-RNN and P-RNN complement and correct each other in the ensemble model to obtain more stable predictions of DI-RNN.

3.2.4 | Network complexity and training time of DI-RNN

The improvement of DI-RNN prediction ability is at the expense of training more parameters, where more parameters mean higher network complexity.³⁵ The number of DI-RNN parameters (373) is twice as much as that of S-RNN parameters (181), not affected by input structure parameters. Fortunately, because basic models are independent of each other that can be trained in parallel, training time is still feasible considering negligible microsecond prediction time of a single value. For instance, the best performing model, that is, DI-RNN(192,4), takes 225 seconds to train, which is relatively short compared to 1800s (30 minutes). This means even if it is required that DI-RNN should be trained from scratch for every single forecast, the training can be completed in the 30-minute gap.

3.2.5 | Applicability of DI-RNN to aggregated load

It is can be seen from Figure 8A that the aggregated load also has significant periodic fluctuation such as 24-hour and 12-hour cycles in frequency domain. The same experiment as individual household is implemented for the aggregated load and the results are shown in Table 3. We present the aggregated load forecasting profiles from 5 February 2014 to 11 February 2014 in Figure 9.

In terms of time domain metrics, “Wavelet+ARIMA” is still the worst performer. MAE and EAIP of DI-RNN are averagely promoted by 2.86% and 6.23% respectively, compared with those of all benchmarks except for “Wavelet +ARIMA.” In frequency domain, the SS of DI-RNN is 43.04% better than that of all benchmarks on average. Figure 8B-H shows the spectrums of DI-RNN/(192,4) prediction and all benchmarks prediction results for the aggregated load, respectively. We can see the energies of DI-RNN at remarkable frequency points such as $\frac{1}{24}h^{-1}$ and $\frac{1}{12}h^{-1}$ are the closest to those of the aggregated load. The constant component of DI-RNN spectrum is also the closest to the actual situation. Therefore, it is concluded that the proposed DI-RNN is applicable to the aggregated load as well, that is, the proposed model has excellent prediction performance for both individual load and aggregated load in time and frequency domains.

4 | CONCLUSION

This paper reveals the deficiency of conventional RNN in capturing periodic information of load from the perspective of frequency domain. A novel STLF model named DI-RNN is presented for accurate prediction, which is based on the proposed input structure, that is, dual granularity inputs (DI) structure. In experiments, we verify the enhanced performance of DI-RNN in STLF from three perspectives, which is available for individual load and aggregated load. Finally, the spectra of DI-RNN prediction results confirm that its ability to capture load periodic fluctuations is boosted.

For future work, we will further explore the characteristics of load data and try to reduce the complexity and training time of DI-RNN for more efficient forecasting.

ACKNOWLEDGEMENTS

This work was sponsored by Shanghai Sailing Program, Grant Number 19YF1423700; the National Key Research and Development Program of China, Grant Number 2016YFB0900100; Key Project of Shanghai Science and Technology Committee, Grant Number 18DZ1100303.

REFERENCES

1. Haider HT, See OH, Elmenreich W. A review of residential demand response of smart grid. *Renew Sustain Energy Rev.* 2016;59:166-178.
2. Tsoukalas LH, Gao R. From smart grids to an energy internet: assumptions, architectures and requirements. Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies; 2008:94-98.
3. Wang Y, Xia Q, Kang C. Secondary forecasting based on deviation analysis for short-term load forecasting. *IEEE Trans Power Syst.* 2011; 26(2):500-507.
4. Wang Y, Chen Q, Hong T, Kang C. Review of smart meter data analytics: applications, methodologies, and challenges. *IEEE Trans Smart Grid.* 2019;10(3):3125-3148.
5. 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(1):188-201.
6. Sun M, Wang Y, Teng F, Ye Y, Strbac G, Kang C. Clustering-based residential baseline estimation: a probabilistic perspective. *IEEE Trans Smart Grid.* 2019;10(6):6014-6028.
7. Kwac J, Flora J, Rajagopal R. Household energy consumption segmentation using hourly data. *IEEE Trans Smart Grid.* 2014;5(1): 420-430.
8. Stephen B, Tang X, Harvey PR, Galloway S, Jennett KI. Incorporating practice theory in sub-profile models for short term aggregated residential load forecasting. *IEEE Trans Smart Grid.* 2017;8(4):1591-1598.
9. Ghofrani M, Hassanzadeh M, Etezadi-Amoli M, Fadali MS. Smart meter based short-term load forecasting for residential customers. 2011 North American Power Symposium, Boston; 2011:1-5.
10. Hu X, Ferrera E, Tomasi R, et al. Short-term load forecasting with radial basis functions and singular spectrum analysis for residential electric vehicles recharging control. 2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC); 2015:1783-1788.
11. Abedinia O, Amjadi N, Zareipour H. A new feature selection technique for load and Price forecast of electrical power systems. *IEEE Trans Power Syst.* 2017;32(1):62-74.

12. Shahzadeh A, Khosravi A, Nahavandi S. Improving load forecast accuracy by clustering consumers using smart meter data. 2015 International Joint Conference on Neural Networks (IJCNN); 2015:1-7.
13. Hong T, Wang P, Pahwa A, Gui M, Hsiang SM. Cost of temperature history data uncertainties in short term electric load forecasting. 2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems, Singapore; 2010:212-217.
14. Amarasinghe K, Marino DL, Manic M. Deep neural networks for energy load forecasting. 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), Edinburgh; 2017:1483-1488.
15. Alzate C, Sinn M. Improved electricity load forecasting via kernel spectral clustering of smart meters. 2013 IEEE 13th International Conference on Data Mining, Dallas, TX; 2013:943-948.
16. Chen Y et al. Short-term load forecasting: similar day-based wavelet neural networks. *IEEE Trans Power Syst*. 2010;25(1):322-330.
17. Pirbazari AM, Chakravorty A, Rong C. Evaluating feature selection methods for short-term load forecasting. 2019 IEEE International Conference on Big Data and Smart Computing (BigComp), Kyoto, Japan; 2019:1-8.
18. Li S, Wang P, Goel L. A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection. *IEEE Trans Power Syst*. 2016;31(3):1788-1798.
19. Wang P, Liu B, Hong T. Electric load forecasting with recency effect: a big data approach. *Int J Forecast*. 2016;32(3):585-597.
20. Cao X, Dong S, Wu Z, et al. A data-driven hybrid optimization model for short-term residential load forecasting. 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing. 2015:283-287.
21. Taylor JW. An evaluation of methods for very short-term load forecasting using minute-by-minute British data. *Int J Forecast*. 2008;24(4):645-658.
22. Gers FA, Schmidhuber J, Cummin F. Learning to forget: continual prediction with lstm. *Neural Comput*. 2000;12(10):2451-2471.
23. Kong W, Dong ZY, Jia Y, Hill DJ, Xu Y, Zhang Y. Short-term residential load forecasting based on lstm recurrent neural network. *IEEE Trans Smart Grid*. 2019;10(1):841-851.
24. Zheng H, Yuan J, Chen L. Short-term load forecasting using emd lstm neural networks with a xgboost algorithm for feature importance evaluation. *Energies*. 2017;10(8):1168-1188.
25. Sehovac L, Grolinger K. Deep learning for load forecasting: sequence to sequence recurrent neural networks with attention. *IEEE Access*. 2020;8:36411-36426.
26. Shi H, Xu M, Li R. Deep learning for household load forecasting—a novel pooling deep rnn. *IEEE Trans Smart Grid*. 2018;9(5):5271-5280.
27. Department of Industry, Innovation and Science, Australian Government. Electricity Consumption Benchmarks. 2014. <https://data.gov.au/dataset/electricity-consumption-benchmarks>
28. Gardner M, Dorling S. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmos Environ*. 1998;32:2627-2636.
29. Schruben L. Confidence interval estimation using standardized time series. *Oper Res*. 1983;31(6):1090-1108.
30. Kingma DP, Adam JB. A method for stochastic optimization. arXiv e-prints. 2014; p. arXiv:1412.6980.
31. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Machine Learn Res*. 2014;15:1929-1958.
32. Huo J, Shi T, Chang J. Comparison of Random Forest and SVM for electrical short-term load forecast with different data sources. 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing. 2016:1077-1080.
33. Yu X et al. Load forecasting based on smart meter data and gradient boosting decision tree. 2019 Chinese Automation Congress (CAC), Hangzhou, China. 2019:4438-4442.
34. Lee C-M, Ko C-N. Short-term load forecasting using lifting scheme and ARIMA models. *Exp Syst Appl*. 2011;38(5):5902-5911.
35. Sum J. Comparative study on various pruning algorithms for rnn. i. Complexity analysis. *Proc Int Conf Machine Learn Cybernet*. 2002;4: 2225-2230.

How to cite this article: Yuan L, Ma J, Gu J, Wen H, Jin Z. Featuring periodic correlations via dual granularity inputs structured RNNs ensemble load forecaster. *Int Trans Electr Energ Syst*. 2020;30:e12571. <https://doi.org/10.1002/2050-7038.12571>