

# The Investigation of Application of Machine Learning into Multi-Level Network Load Forecasting

**Yang Haolin**

*George Washington University, USA*

**ABSTRACT.** *In this report, three different deep learning algorithms: RNN, RGU and CNN are applied for time series forecasting into the distributed level system. The necessary dataset cleaning is to make sure the failure outcome due to the raw data. To determine the best methods for each network load level, the machine learning models are also compared with the MSE and MAPE of naive forecasting model to verify its feasibility. With the enhancement of the sum customer loads, the forecasting result become better and accuracy.*

**KEYWORDS:** *Smartgrid ; RNN ; CNN ; RGU; Naive Forecasting*

## 1. Introduction

System operators commonly use aggregate load and PV generation forecasting techniques to effectively manage the electric grid operation. Facing more challenges from nonlinear appliance and electric device, the forecasting technology are playing more essential roles in the demand response project, especially in the era where the distributed generations and energy storage system are heavily customer-oriented. Unlike electric load at the system level, domestic power consumptions are often with high volatility, which makes meter level load forecasting for a single user extremely challengeable

## 2. Data Process

### 2.1 Residential data

The apartment dataset contains electric usage data for 114 single-family apartments for the period 2014-2016. The device collecting the data is a sort of smart meter (eGauage) and all the apartments are in the Western part of Massachusetts state. This dataset identified residential customers with separate meters that record the usage of individual apartment. The time format of the

Residential Dataset is recorded into the intervals of every minutes and are transferred into the timestamp format while being input into the machine learning model. The observation of the dataset is as shown into the table 1 and see more detail: <http://traces.cs.umass.edu/index.php/Smart/Smart>

*Table 1 Information of the 114 Apartments Dataset*

Items	Details
Number of customers	114.
Date: 1 January 2016 to the 15 December 2016	349 days.
A single day: T	24 h from midnight
Number of daily intervals: s	24*60=1440.
Interval period:	1min.
Number of Weather Parameter	14.
Number of Non-linear Lable	2.
Number of daily Weather intervals: s	24*4*96
Interval Weather Period:	15min.

## 2.2 Clean dataset

The main method of cleaning data is referred [5]. In principle, the dataset that is referred consists of load usage and time data for each of 114 customers at 1-min resolution. In practice, a clean dataset is required to ensure the robustness and feasibility of the results of the forecasting model. However, leading to anomalous measurement, several factors exist will be identified and discussed in this section.

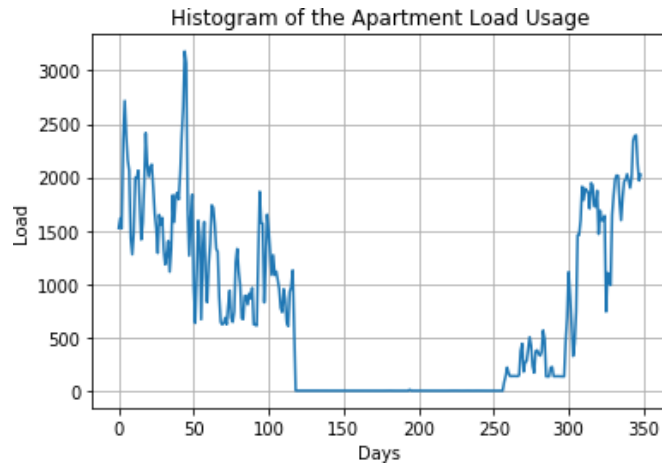
**Repeated Record and Missing Record:** There is no missing file of the 114-apartment dataset. However, while going through all the data file, within every file of record, the repeated data and missing gap are found at the same time point. There is a 60-point-gap existing between the 1:59 AM and 3:00 AM at March 2016. the repeated data occurs at the data of 6 November 2016 and after 01:00 AM. The daylight save condition contribute to such problem. Since the output is collected in EST.

**Low Load Record:** The apartments would be removed, with the anomalous records on long-lasting low power in one-year period that potentially raised when the customer is undergoing for travel and vacation. Such dataset is meaningless to input the forecasting model. More specifically, the apartments with max daily usage less than 50 W and lasting 40 days are removed. Luckily, only on

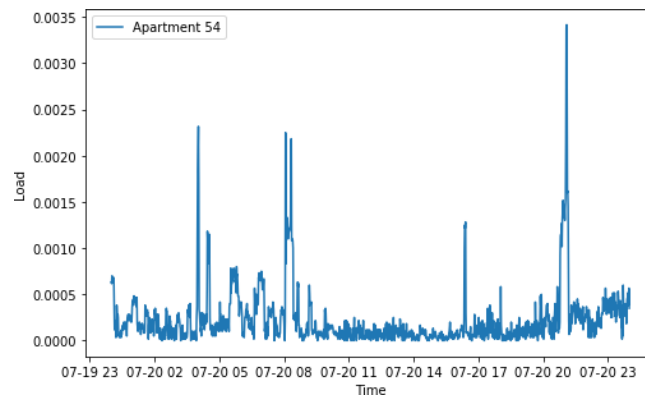
The ID of the apartment is 54. The Figure 1 shows the load curve of the related apartment of the whole year.

**High Spike Records:** At the time point of 2016/3/27 10:21:00, there are 70 apartments consisting a spike whose the magnitude is over 20. It is not acceptable to aboard all the files with the huge spike. Even a little change of the raw data may bring the bias to model result. The solution is to interpolate the data. The average of the load of T-1 and T+1 replaces the spike value. The solution is to interpolate the

data. The average of the load of T-1 and T+1 replaces the spike value because the timestamps is consistence, which will not affect the accuracy of the forecasting



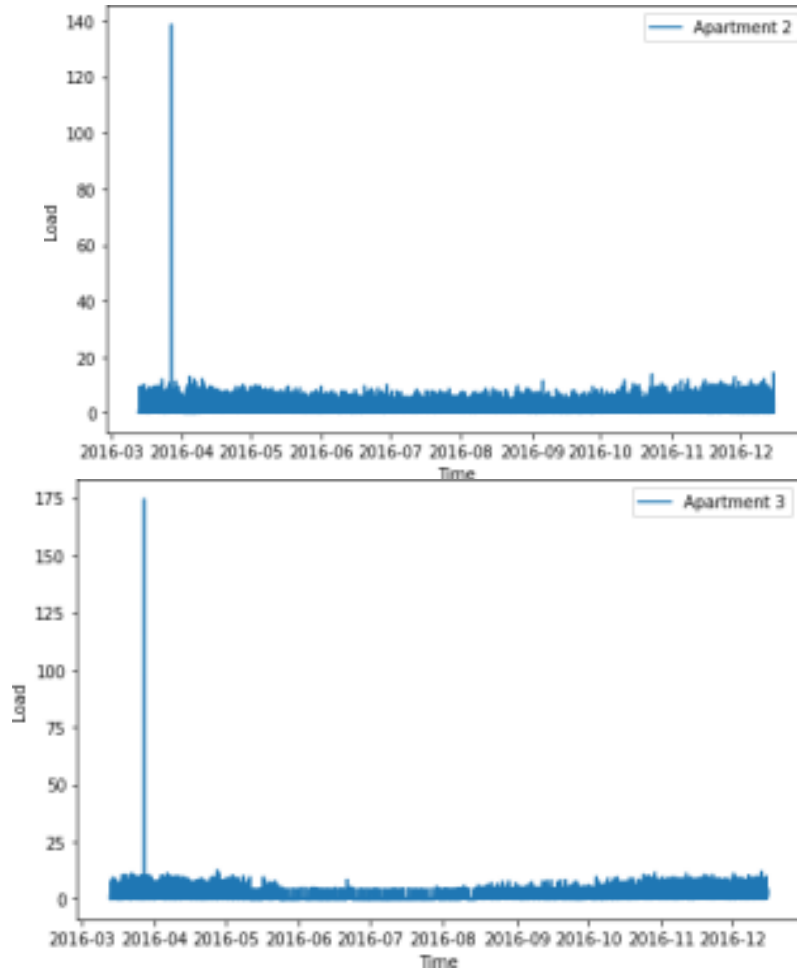
*Fig. 1. The Apartment 54 Load Curve for overall year*



*Fig. 2. The Apartment 54 Load Curve for one day*

### 3. Aggregated Analysis

In Figure 5 shown as following. According to the time index on each of the 314 days, the residential demand of 111



*Fig. 3. The Spike in Apartment 2 and Apartment 3*

Table 2. The Apartments with The Huge Spike

ID :							
2	3	5	6	7	8	9	10
11	12	14	16	18	19	21	22
23	24	25	29	30	31	32	35
36	38	40	41	42	43	45	46
50	53	54	55	56	57	58	59
62	65	68	69	70	71	72	75
78	80	81	83	84	85	88	94
96	97	98	99	101	102	103	104
106	107	108	110	111	112		

Customers in the cleaning dataset are summed and presented in an aggregated form.

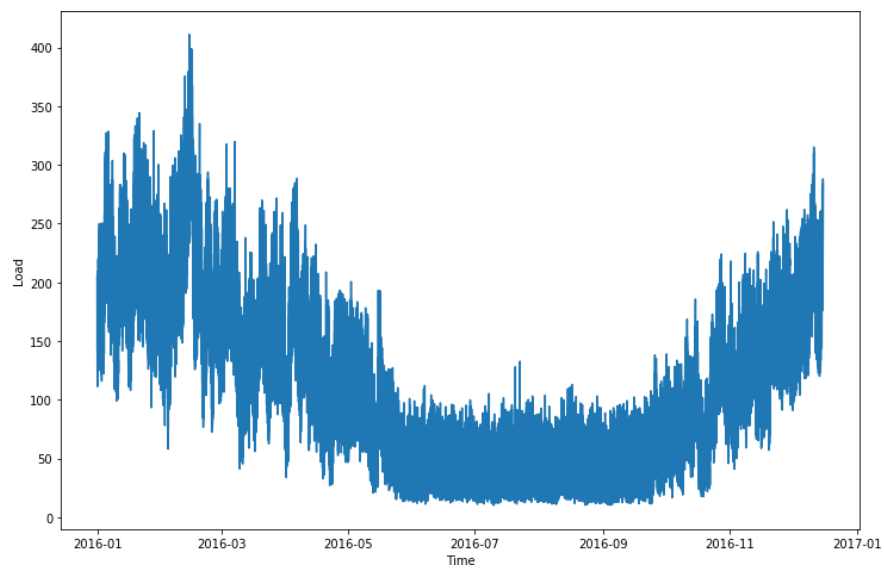


Fig. 4. Aggregated Residential Demand Load with Cleaning Dataset

As we can observe, the peak aggregated demand is about 402 during the winter. During the winter of this year, the peak aggregated demand is greater than the summer aggregated peak of 2016. Followed by a period of stabilizing from the June to October, the demand starts to fall since the April and tend to an enhancement beginning at October. Unlike the residential demand of the winter, the customer

consumption in the summer do not raise to a peak in the night. This may have related to the local geological factory, for example, people prefer to open the heat when cold winter is coming and close the AC while embracing the breezy summer.

Next, 4 arbitrary of customers in the cleaning dataset are selected to investigate the daily residential load variability. With an intention to emphasis on the variability in customer- specific load.

By observation, the pattern of usage of load in these 4 apart- ments are different. As such, no visual confirmation regarding a strong relationship in energy consumption on consecutive weekdays is evident that can be observed in these profiles

#### 4. PROPOSED FORECASTING METHOD

In this section, three deep learning method that are applied into the forecasting of network load are illustrated. The method includes RNN, CNN and GRU.

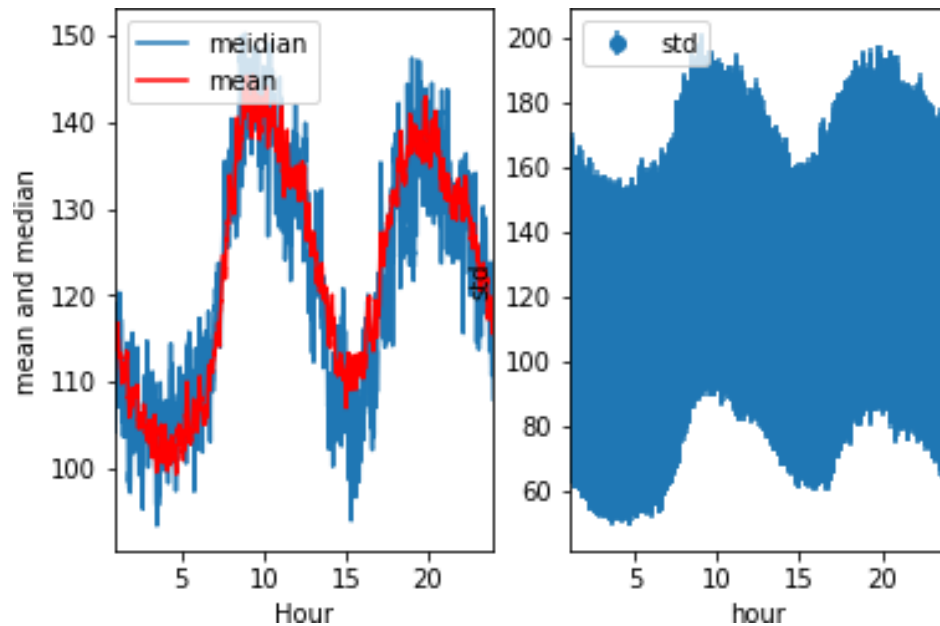


Fig. 5. Aggregated Winter Residential Demand Load with Cleaning Dataset

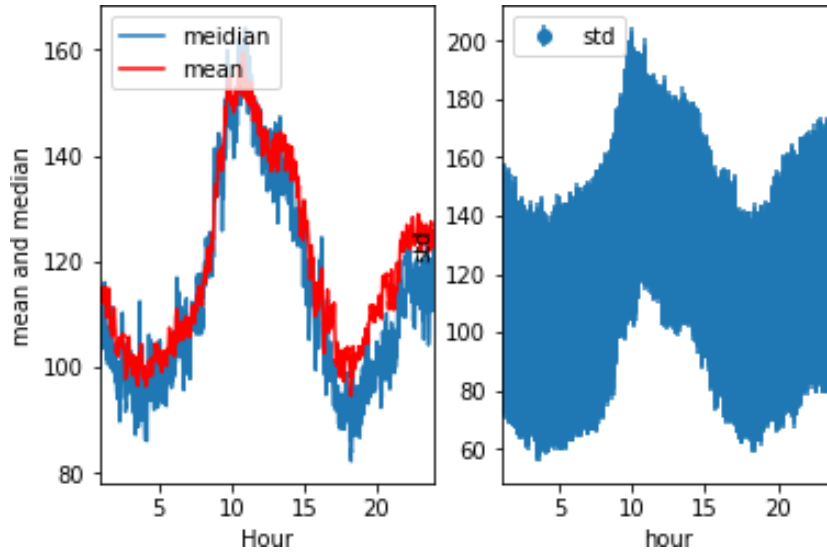


Fig. 6. Aggregated Summer Residential Demand Load with Cleaning Dataset

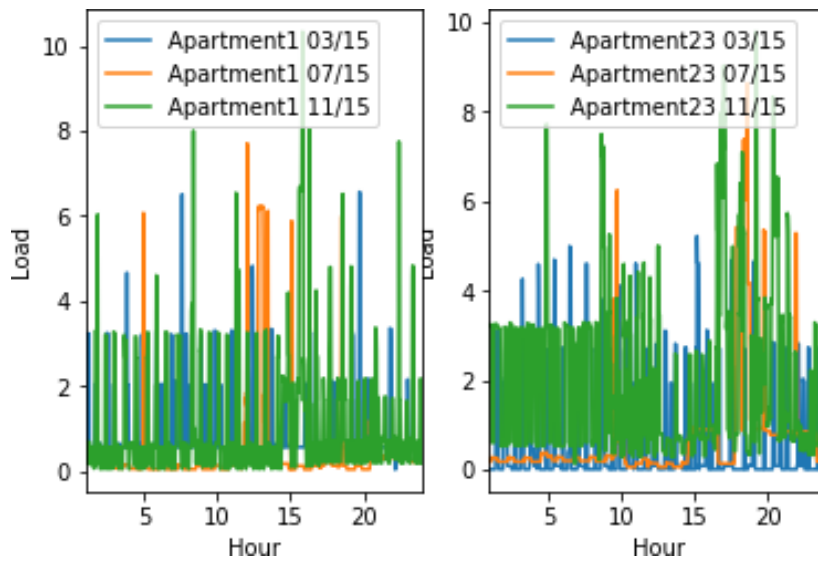


Fig. 7. Aggregated Residential Demand Load of Apt 1 and Apt 23

website: <https://medium.com/@huangkh19951228/predicting-cryptocurrency-price-with-tensorflow-and-keras-e1674b0dc58a>

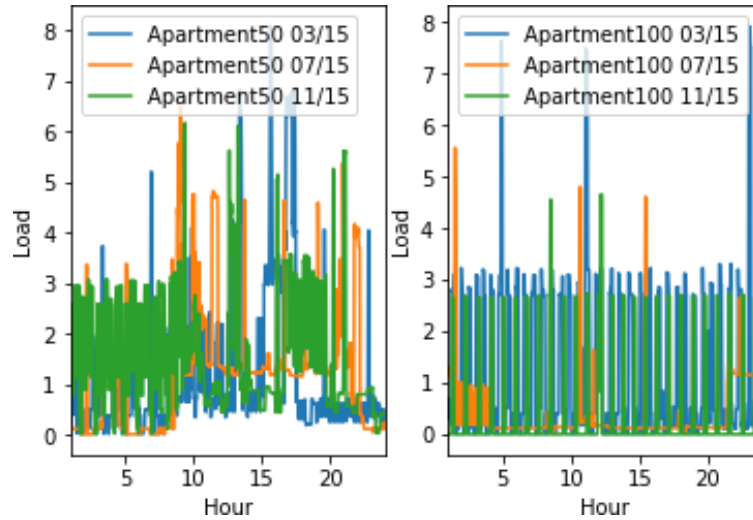


Fig. 8. Aggregated Residential Demand Load of Apt 50 and Apt 100

Two the version of different RNN are applied. The simple GRU (Gated Recurrent Units) as another variation of RNN, is also applied to be supplement of the method test. The LSTM, being short for Long short-term memory, is applied to the forecasting of aggregation analysis of the different levels of load, because of simple RNN do not equipped with satisfied capacity of prediction.

Layer (type)	Output Shape
lstm_1 (LSTM)	(None, 8, 50)
lstm_2 (LSTM)	(None, 50)
dropout_26 (Dropout)	(None, 50)
dense_26 (Dense)	(None, 1)
leaky_re_lu_1 (LeakyReLU)	(None, 1)

Fig. 9. RNN Framework



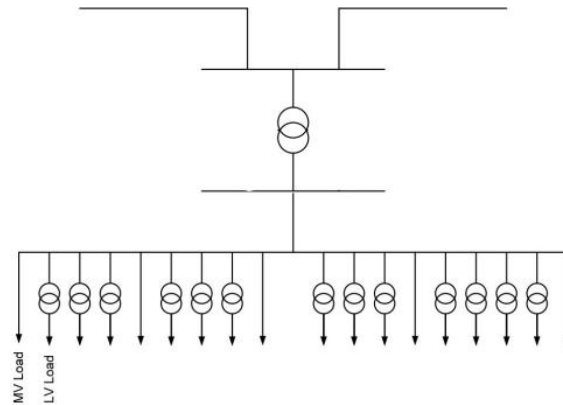
Layer (type)	Output Shape
gru_1 (GRU)	(None, 50)
activation_25 (Activation)	(None, 50)
dropout_25 (Dropout)	(None, 50)
dense_25 (Dense)	(None, 1)
activation_26 (Activation)	(None, 1)

Fig. 10. GRU Framework

The convolutional neural network (CNN, or ConvNet), one of most classic type artificial neural networks has successfully been applied to analyzing visual imagery.

Layer (type)	Output Shape
conv1d_163 (Conv1D)	(None, 15, 8)
max_pooling1d_163 (MaxPoolin	(None, 7, 8)
conv1d_164 (Conv1D)	(None, 6, 8)
max_pooling1d_164 (MaxPoolin	(None, 3, 8)
dropout_185 (Dropout)	(None, 3, 8)
flatten_82 (Flatten)	(None, 24)
dense_185 (Dense)	(None, 250)
dropout_186 (Dropout)	(None, 250)
activation_163 (Activation)	(None, 250)
dense_186 (Dense)	(None, 2)
activation_164 (Activation)	(None, 2)

Fig. 11. CNN Framework



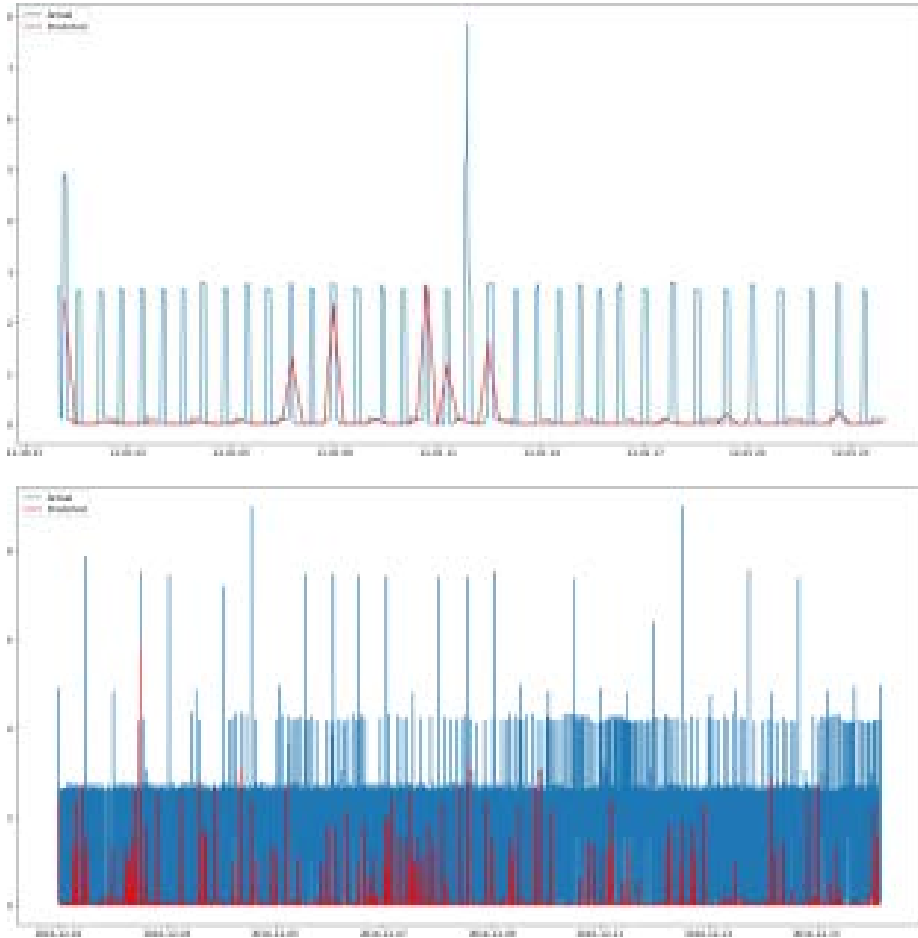
*Fig. 12. Distributed Level System*

#### **4. Iistributed Level Power System**

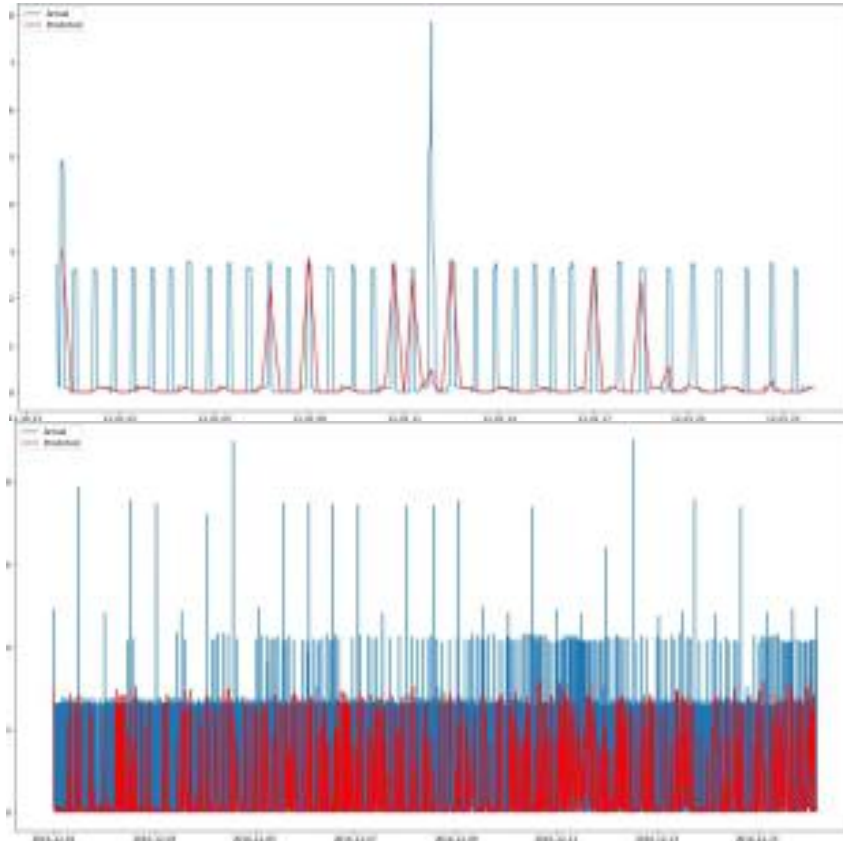
In our model, according to different voltage level with connecting to transformers, the 3 types of aggregated nodes are defined. The first type of node, called Residential Aggregated Node (RAN), represents the individual residential load level. The second types node, named after Transformer Aggregated Node(TAN), connect to the high voltage side of customer transformers and low voltage side of the terminal transformer. Each of such type of node aggregates 5 customers' load demand. Finally, the third type is the node that connected to the transmission level bus, which aggregates all the customers' load demand and is called Terminal Transformer Aggregated Nodes (TTAN).

#### **6. Experiments**

The three deep learning models were performed with com- paration with the naive forecasting to verify their feasibility. The dataset is the cleaning dataset in former section. The three types of aggregated nodes were forecasting and reported with lose values and forecasting profiles. Finally, the best method for each aggregated node was determined considering the accuracy.



*Fig. 13. The CNN Forecasting results of the Apt 30 and Apt 90 upper two:Apt 30 with resolution of one day and one month lower two:Apt 90 with resolution of one day and one month*



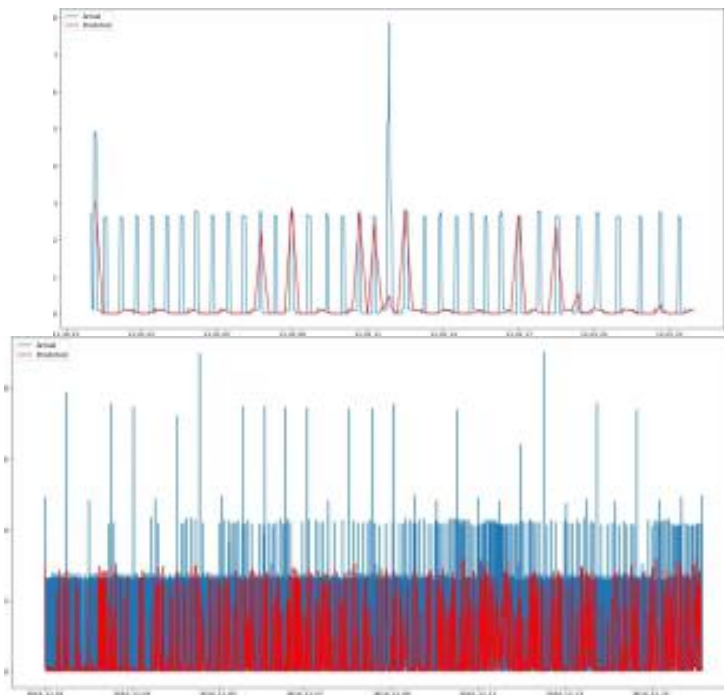
*Fig. 14. The GRU Forecasting results of the Apt 30 and Apt 90 upper two:Apt 30 with resolution of one day and one month lower two:Apt 90 with resolution of one day and one month*

## 7. RESULTS

In this section, we use mean absolute percentage error (MAPE) to compare the performance of three deep learning methods with naive forecasting. The MAPE, the most common measure of forecast error, however, do not perform well when there are extremes to the data and zeros, which is not avoidable especially in the individual dataset. It can conclude the feasibility and valid the necessary of the appliance of deep learning into time series network load forecasting. To make the profiles easy to observe, the mse will not show on the figures because of its over huge value.

The results of selected 100 Residential Aggregated nodes are performed. The forecasting result of Apt 30 and Apt 90 are selected as the examples to illustrate the performance of the three deep learning methods.

For MAPE part, most of the apartment of three deep learns method has a smaller error rate then naive forecasting one. RNN give the average MAPE which is 60.270 While CNN and GRU are separately 39.66 and 87.96. The average outcome of the naive forecasting is 322.49. There is an extreme example existing, which is Apt 54 and its MAPE is around 5435. This is mainly because that there is numerous extremely small value into the dataset. Although the GRU have a relatively better result, the RNN's outcome is still acceptable.



*Fig. 15. The RNN Forecasting results of the Apt 30 and Apt 90 upper two: Apt 30 with resolution of one day and one month lower two: Apt 90 with resolution of one day and one month*

Represent the forecasting results of load spikes and off-peak difference.

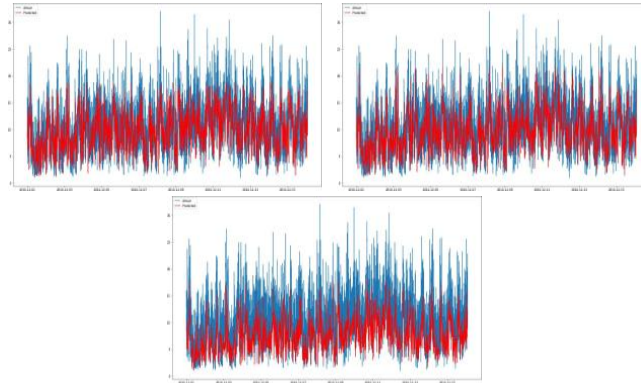


Fig. 16. The Forecasting of 15th node of the Transformer Aggregated nodes  
upper:GRU and RNN lower:CNN

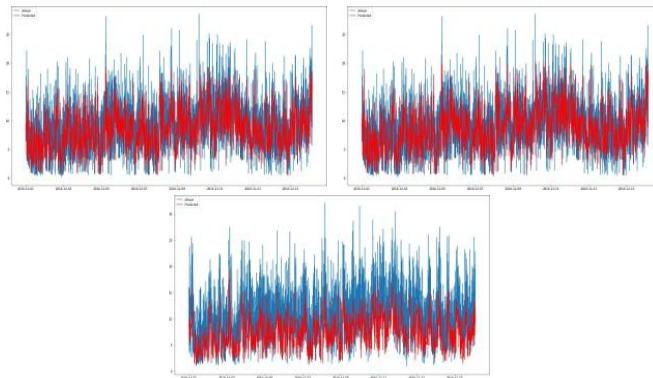


Fig. 17. The Forecasting of 20th node of the Transformer Aggregated nodes  
upper:GRU and RNN lower:CNN

With node 17, the RNN do not have good performance whose MAPE is larger than the naive forecasting. The GRU and CNN defeats all the naive forecasting results and performs steadily among all the nodes. The off-bottom difference description is much more accuracy and response to the load spikes is still ideal enough, though it become much better. The 'Waterloo' of the CNN is a little out of our expectation – the overall results only can predict the tendency of the load and do not response to the spike anymore.

The outcome of the MAPE of three methods are ideal, even though the largest error of the CNN is smaller than the naive forecasting. Compared with different epochs, the RGU and RNN gained the lowest MAPE loss and finally stable at the point of 12 The forecasting results of CNN still cannot

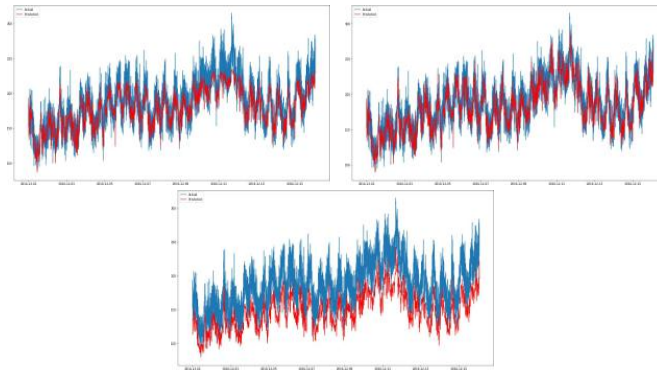


Fig. 18. The Forecasting the Terminal Aggregated nodes upper:GRU and RNN lower:CNN

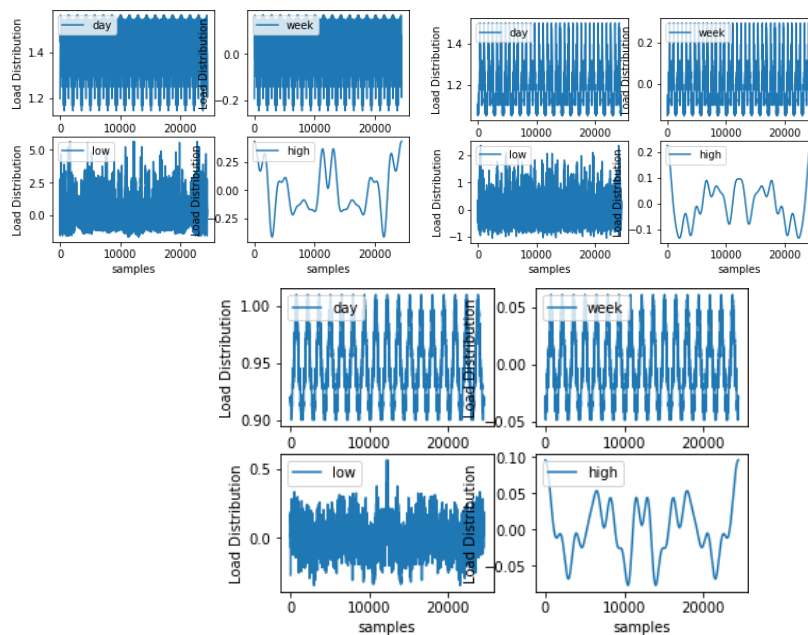


Fig. 19. The Frequency Analysis for 3 Type of Nodes upper left and right:RAN RTN and lower:RTNN

## 8. OVERALL RESULT

Firstly, in general, the naive forecasting method cannot compete with the three diverse deep learning methods. Secondly, we can notice that both the MAPE

become smaller and the forecasting results of the RNN and GRU become better. In the residential aggregated level, the RGU have the lowest MAPE and RNN gain the lowest MAPE values both in the terminal aggregated nodes and transformer aggregated nodes. For every different level, the GRU and RNN only have a little distance of accuracy. Considering the compromise of the a little reducing MAPE and fast speed, the GRU is the most practice method and the RNN is the most exact method in the level of transformer and terminal.

*Table 3. The Outcome for Different Nodes*

	RNN	GRU	CNN	Naive
RAN(MAPE)	60.27	39.36	87.97	322.94
TAN(MAPE)	22.66	24.05	643.40	74.25
TTAN(MAPE)	12.99	13.02	14.90	20.49
Time(s/Epoch)	19	9		9

## 9. Discussion

The network load is a type of strong periodic time series , so we can analyses the network load series within Frequency domain. For this model, the Fourier decomposition is applied to decompose original time series into four diverse time series components.

$$P(t) = a + D(t) + W(t) + L(t) + H(t) \quad (1)$$

The a+D(t) is the daily periodic component whose period is

24 hours(1440 samples). The W(t) is the weekly periodic component whose period is 1 week(1440\*7 samples). The L(t) is the low frequency remain component whose period is larger than 24 hours. The H(t) is the high frequency remain component whose period is smaller than 24 hours.

The results show the nodes with fewer aggregated customers load have more drastic change and are effect more from weather factors and random elements. The difficulty of the selection of the relatively exact forecasting model is enhanced, and so do the challenge of the forecasting.

## References

- [1] Weicong Kong, Zhao Yang Dong, David J. Hill, Fengji Luo and Yan Xu (2018). Short-Term Residential Load Forecasting Based on Resident Behaviour Learning, IEEE Transactions On Power Systems, Vol. 33, No. 1, pp.1



- [2] Liang Xiao, Xingyu Xiao, Canhuang Dai, Mugen Peng, Lichun Wang and H. Vincent Poor(2018). Reinforcement Learning-based Energy Trading for Microgrids, In Press..
- [3] Huaizhi Wang, Jiaqi Ruan, Guibin Wang, Bin Zhou, Yitao Liu, Xueqian Fu, Jianchun Peng(2018). Deep Learning Based Interval State Estimation of AC Smart Grids against Sparse Cyber Attacks , In Press.
- [4] L.M. Borges, F.J. Velez and A.S. Lebres (2014). Survey on the characterization and classification of wireless sensor network applications. Communications Surveys & Tutorials IEEE, vol.16, no.4, p.1860-1890.