Load Forecasting

Using Echo State Network

by

Shahrzad Baraeinezhad

A report submitted in partial fulfilment of the requirements for the Data Engineering Advanced Project II

MSc in Data Engineering

Instructors and Supervisors
Prof. Adalbert Wilhem

Date of Submission: January 7, 2018

Focus Area: Mobility, School of Engineering and Science

Acknowledgements

I would like to thank my supervisor, Prof. Adalbert Wilhem, for giving me this opportunity to conduct this advanced project in the field of Data Engineering. This has been a great experience so far, and I am looking forward to continue working in the field of Data Engineering as my future career. I am thankful of my supervisor for his support thoughout this project.

I would like to thank all my friends who have helped me. In particular, I am indebted to Mr. Sabin Bhandari, a computer science student, for giving me technical support in debugging the code and giving suggestion, as well as, his assistance in looking through my report as another set of eyes.

Abstract

The purpose of this project is to predict the power load by using Echo State Network (ESN). After going through many machine learning methods, I found ESN to be one of the most efficient method for learning nonlinear systems. Since, the data that I used for the project is a time series, I found ESN to be well suited for the task. As the population grows, there is an increase in energy demand made on existing electricity supply. The prediction of energy is essential to design and size suitable renewable energy systems and energy storage. It has many applications including energy purchasing and generation, load switching, contract evaluation, and infrastructure development. That is why there is a need of precise prediction model of power load.

Contents

List of Figures

List of Tables

Acronym

ESN Echo State Network

NSLS Net System Load Shape

SSS Standard Supply Service

THESL Toronto Hydro Electric System Limited

KWH Kilo Watt Hour

RNN Recurrent Neural Network

LSM Liquid State Machines

NRMSE Normalized Root-Mean-Square Error

MAPE Mean Absolute Percent Error

SCM State Collection Matrix

MAE Mean Absolute Error

EP Evolutionary Programming

SVM State Vector Machine

1 Introduction

1.1 Load Prediction

Electricity is an important source of energy that is consumed in the household. It can convert into any other forms of enery suh as heat, light, sound, magnetic, etc, thus it plays an important role in the day to day life of human beings. Not only in household, but also in other sector like industry, power is a driving fator for the productivity. Thus, the load forecasting plays an essential role in the planning of electricity industry and the operation of electric power systems. It is important to devise an accurate forecast model to increase the reliability of power supply and delivery system. It leads to substantial savings in operating and maintenance costs and correct decisions for future development. Electricity demand is assessed by accumulating the consumption periodically; it is almost considered for hourly, daily, weekly, monthly, and yearly periods [4].

Load prediction can be divided into three categories: (i) short-term forecasts which are usually from one hour to one week, (ii) medium forecasts which are usually from a week to a year, and (iii) long-term forecasts which are longer than a year. The forecasts for different time horizons are important for different operations within a utility company.

1.2 Dataset Description

The Net System Load Shape (NSLS) defines the profile of consumers who will receive smart meters. It is an aggregate profile of residential and small commercial profiles. According to Toronto Hydro, all other Standard Supply Service (SSS) customers (customers that have not signed a retail contract) are charged based on Toronto Hydro Electric System Limited (THESL) NSLS. The NSLS is specific to the local utility, and is calculated by taking the overall THESL profile and deducting the sum of all interval meters and fixed load profiles (such as street lighting). Customers with interval meters (specialized meters which register hourly energy use) are charged based on their hourly load profile.

The THESL NSLS data for the time period of August 2016 - October 2017 is obtained from [9] https://www.torontohydro.com. The summary of data set is given in 4 where for each meter there is time in hours and power consumption data associated with it inKilo Watt Hour (KWH).

1.3 Echo State Network

ESN is a Recurrent Neural Network (RNN) architecture which is made up of two components: reserviour which has a recurrent topology of nonlinear elements and a readout which is a memoryless linear network. The main idea is to drive a random, large, fixed recurrent neural network with the input signal, thereby inducing in each neuron within this ”reservoir” network a nonlinear response signal, and combine a desired output signal by a trainable linear combination of all of these response signals [3]. The reservoir state is called ”echo states” and they contain information about the history of input patterns. It is a special kind of RNNs which is comparatively cheaper and faster method of supervised learning [7]. The fundamental concept of ESN is shared with Liquid State Machines (LSM), which were developed independently from and simultaneously with ESNs by [8].

1.4 Objective

The main objective of this project is to predict the power consumption and compile the results with essentially includes the error and level of accuracy of the method. For this project, I would also analyze how well ESN performs in this data set.

2 State of the art

Many different statistical and artificial intelligence techniques have been developed and used for short-term load forecasting. By using different influence factors such as holidays, average loads and weathers, [10] presented several regression models for the next day peak forecasting. [2] describe a practical real-time implementation of ARIMAX models for load forecasting. Offline testing and online operation has consistently shown satisfactory performance with the Mean Absolute Error (MAE) mostly less than 2% for a less than 24-hour ahead forecast and less than 2.5% for a less than 168-hour ahead forecast. [12] used evolutionary programming (EP) approach to identify the ARMAX model parameters for one day to one week ahead hourly load demand forecast.

Fuzzy logic has been applied to the problem of peak load forecasting for the Greek power system by [5], where different fuzzy expert systems have been assigned to each season of the year.The resulting fuzzy expert system constitute a highly effective forecaster of peak loads, attaining a yearly average forecast error of 2.45% that varies between 1.32% and 3.62%. [1] proposed a State Vector Machine (SVM) model to predict daily maximum load of the next 31 days, aiming at mid-term load forecasting. Their program was the winning entry of the competition organized by the EUNITE network in 2001.

3 Theoretical Background

ESN is composed of neural networks with (i) $\mathbf{u}(n)$ K-dimensional input, (ii) $\mathbf{x}(n)$ N-dimensional reservoirs, and (iii) $\mathbf{y}(n)$ L-dimensional output.

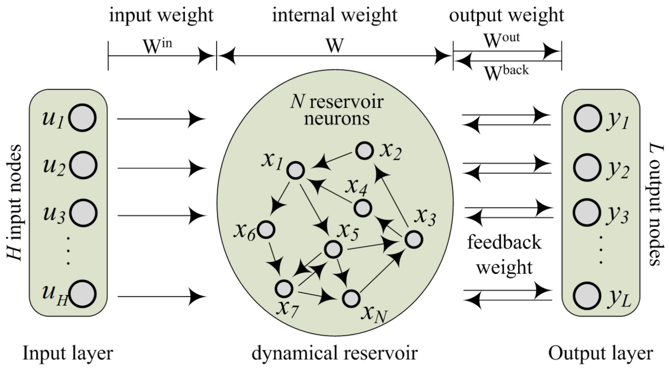

Figure 1: The basic schema of an ESN from [6].

W is the $N \times N$ reservoir weight matrix, $\mathbf{W}^{in}$ is the $N \times K$ input weight matrix, and **b** is an N-dimensional bias vector. f is a sigmoid function (tanh in our case). The state update equation for the activation of internal units which is

$$\mathbf{x}(n+1) = f(\mathbf{W}\mathbf{x}(n) + \mathbf{W}^{\mathbf{in}}\mathbf{u}(n+1) + \mathbf{b}) \tag{1}$$

The extended system state is obtained by the concatenation of the reservoir and input states at time n which is

$$\mathbf{z}(n) = [\mathbf{x}(n); \mathbf{u}(n)]. \tag{2}$$

When the ESN is driven with input, the extended system is obtained. It is given as:

$$\mathbf{y}(n) = \mathbf{W}^{out}\mathbf{z}(n) \tag{3}$$

$\mathbf{W}^{out}$ is a $L \times (K+N)$-dimensional matrix of output weights.
Then, these states are filed row-wise into a State Collection Matrix (SCM) **S** of size $(n_{max} \times (N+K))$. The desired outputs **d**(n) are kept row wise into a teacher output collection matrix **D** of size $(n_{max} \times L)$ [3]. The output weights $\mathbf{W}^{out}$ is calculated as:

$$\mathbf{W}^{out} = (\mathbf{S}^t\mathbf{S} + \alpha\mathbf{I})^{-1}\mathbf{S}\mathbf{D}^t \tag{4}$$

where $\mathbf{S}^t\mathbf{S}$ is the correlation matrix with **S** being SCM , **D** is the desired output matrix and α is the regularization coefficient.

The spectral radius of the reservoir weight matrix is the maximal absolute eigenvalue of this matrix and scales the width of the distribution of nonzero elements of **W** [3]. It is applied to initial random weight matrix.

4 Error Calculation

The error is calculated in two ways. Normalized Root-Mean-Square Error (NRMSE) is calculated between desired output weights $\mathbf{d}(n)$ and output signal **y**(n) as:

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{N}((\mathbf{y}(i) - \mathbf{d}(i))^2}{N\hat{\sigma^2}(d)}} \tag{5}$$

where **N** = total output data points and $\hat{\sigma}$ = variance.

Mean Absolute Percent Error (MAPE) is calculated between desired output weights **d**(n) and output signal **y**(n) as:

$$MAPE = \sum_{i=1}^{N} \left| \frac{\mathbf{y}(i) - \mathbf{d}(i)}{\mathbf{y}(i)} \right| \tag{6}$$

5 Methodology

5.1 Visualization of Data

The data set is processed before the machine learning technique is applied. The METER ID is of no use as the same meter is used to calculate the load, so the column is disregarded. Them, the plot between the time and power is taken which is shown in the figure 2.

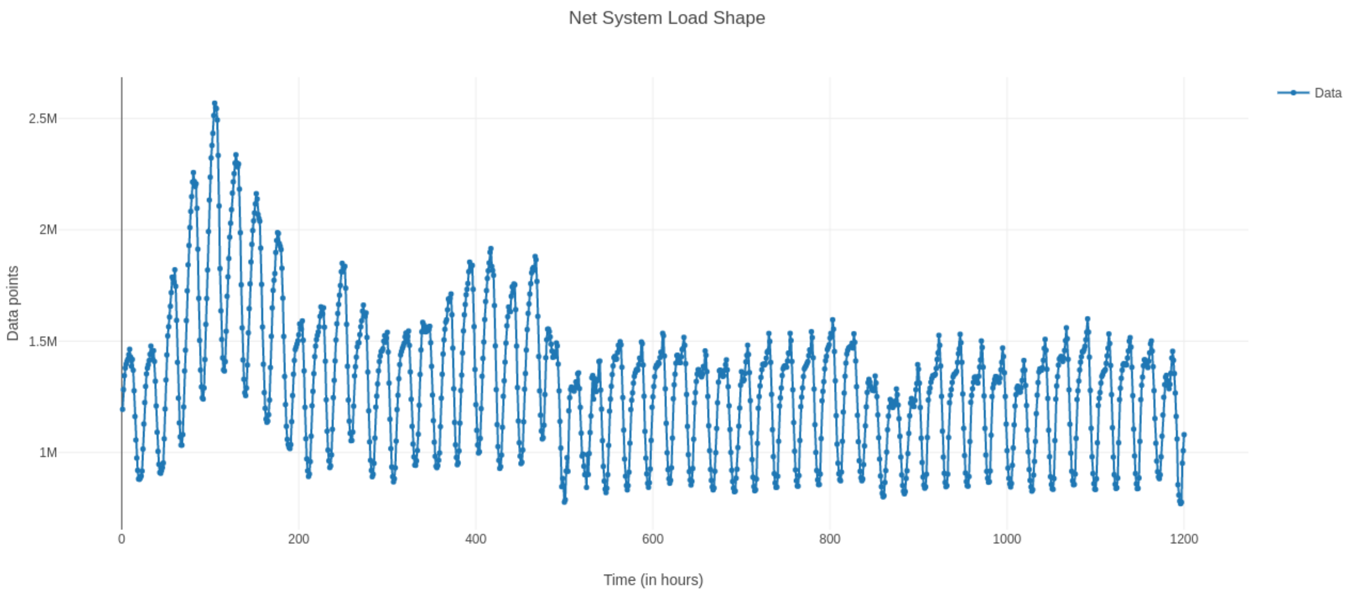

Figure 2: Plot of data points.

5.1.1 Processing of Data

Logarithm function is applied to squeeze the data set into meaningful level. Then to make a standard input for the ESN, the formula is applied as

$$x' = \frac{x - \bar{x}}{\sigma} \tag{7}$$

where (i) x = data,(ii) $\bar{x}$ = mean , (iii) and σ = standard deviation.

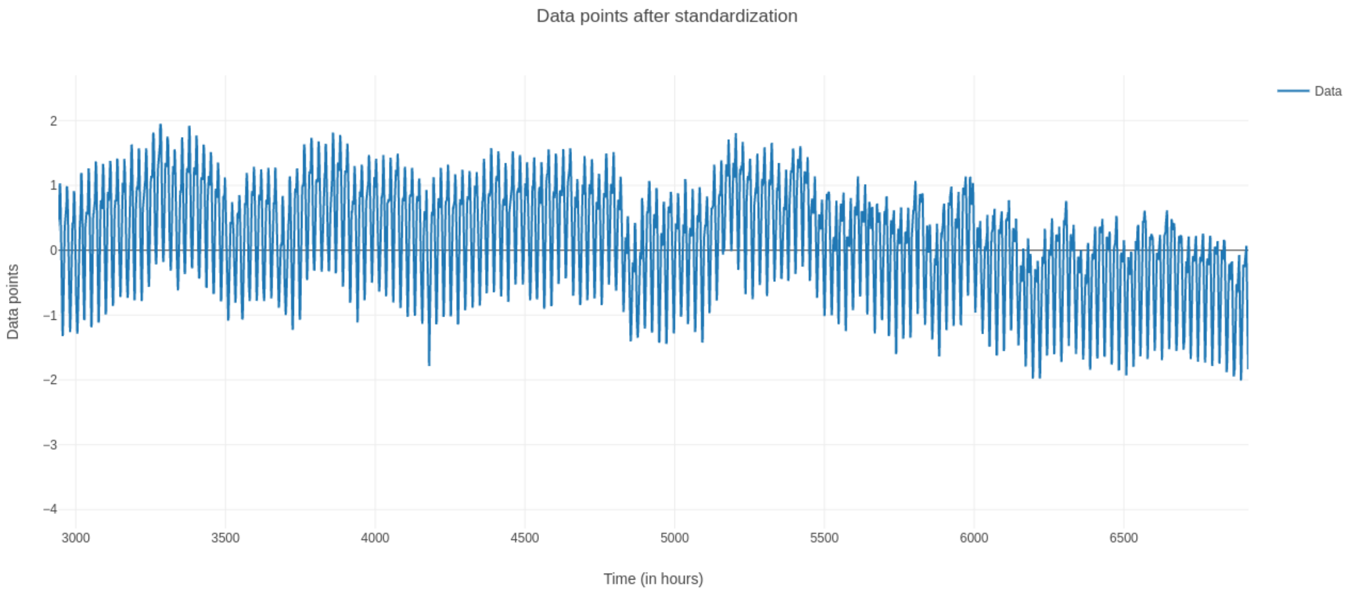

Figure 3: Standarized data points

5.2 Structure of Network

For our prediction task, the minimal form of ESN is sufficient. In my implementation, I took 1 input which is the collected data and 1 output which gives the prediction. The internal units is 1000 sized. The standarized data is introduced to the reservoir with 1000 units. The weights are randomized and normally distributed. The spectral radius, which is used to ensire the echo state property, is calculated as $\rho(\mathbf{W}) = max(\ abs(\ eig(\ \mathbf{W}\ )))$, where in my implementation it is taken to be 1. The weights are scaled respectively by input weight scaling factor $(Winsf)$, reservoir weight scaling factor (Wsf) and bias scaling factor (bsf).

5.3 Training of data set

Out of 10968 data points, I took 8568 data points as input and remaining 2400 data points as testing data (final data kept aside while training). All the data points were already processed. Then the training data points i.e. input is subjected to cross-validation process.

5.3.1 K fold Cross-validation

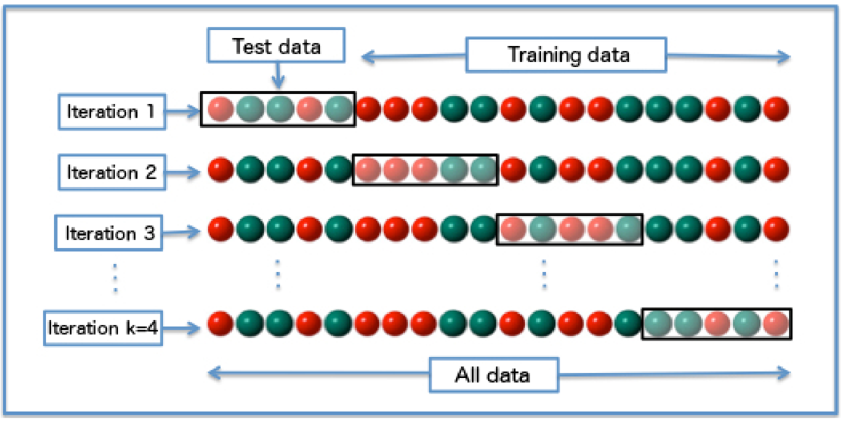

Figure 4: k fold cross validation taken form [11].

The training data points of size 8568 is split into K folds. In my case K is taken as 6. During this phase, the data set runs into K cycles, where K-1 data set is taken as training partition and the remaining 1 fold is taken as validation partition. In each cycle, validation error is calculated and the parameters which gives the optimal solution is taken. The judgement is on the basis of NRMSE error.

5.4 Optimization of the model

The optimizing parameters for this model are:

- Scaling factors: $Winsf$, Wsf and bsf.
- Regularizing cofficient: α
- Reservoir size: N

For each of the parameter, the value is changed one at a time to really see how the model is performing. Change in NRMSE error is the basis for us to judge the performance of the model on the basis of the changes in the parameters. Also, there is a need of cleaning of the initial transient by a certain cut off value. In this case it is taken as 400 by looking at the graph of the output.

5.5 Testing Phase

This is the final step for the prediction task. After the cross-validation, we have the parameters which are optimal for our working model. The final data of size 2400 is taken as input. Then, by keeping all the weights and parameter optimal, the output **y** was calculated by using $\mathbf{W}^{out}$ and the desired output **d** (refer to 2). Then, the NRMSE and MAPE were calculated to check learning capacity of the network.

6 Result

6.1 One hour prediction

Training phase

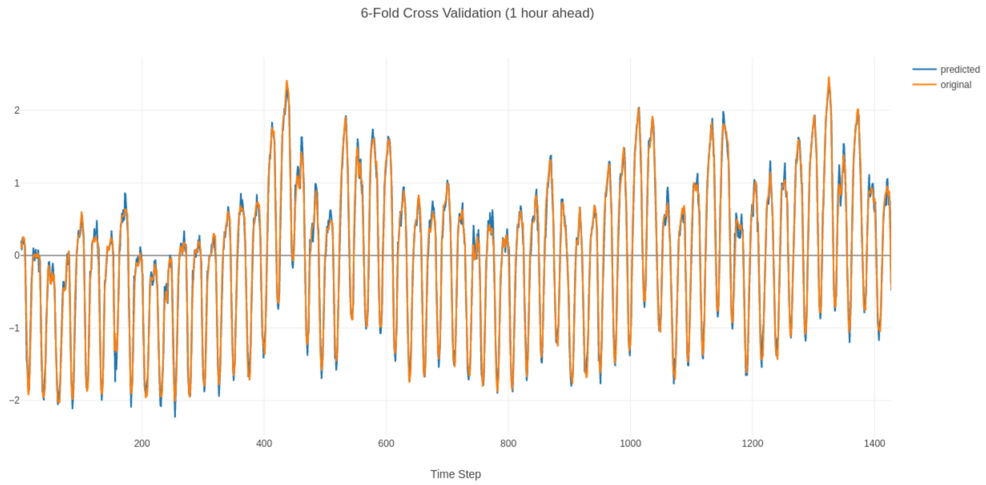

Figure 5: plot between training and validation in cross-validation.

Testing phase

After using the testing datasets, the desired test data and the output test data are plotted after calculation.

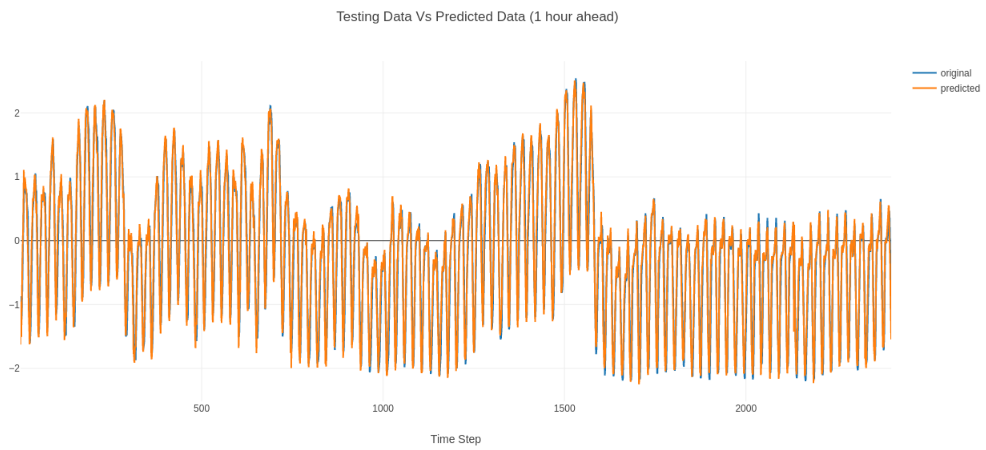

Figure 6: Result after testing the data.

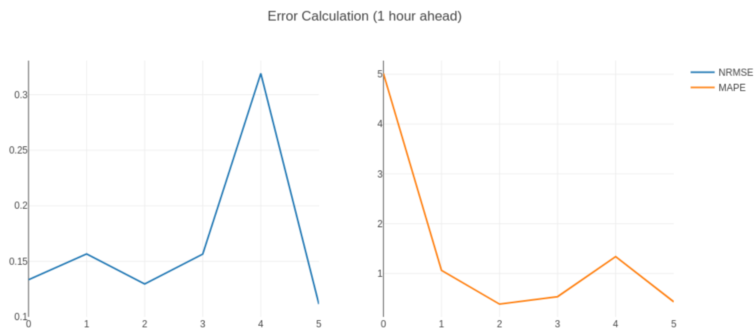

Figure 7: NRMSE at each fold.

The Cross-validation parameter values for the optimal result are: Wsf = 2 , $Winsf$ = 1.5, $bias$ = 0.48 and α = 0.028. During training phase, the error was computed on each validation set. The plot of each validation set and predicted result is shown in figure 5. The resulting error on each validation set is shown in Table 1.

	Partition 1	Partition 2	Partition 3	Partition 4	Partition 5	Partition 6
NRMSE	0.13	0.156	0.129	0.156	0.319	0.11
MAPE	5.0167	1.063	0.385	0.535	1.339	0.432

Table 1: NRMSE and MAPE error in each partition for 1 hour prediction.

The mean NRMSE error is 0.167 and mean MAPE error is 1.46. The plot of NRMSE and MAPE of each partition are shown in figure 7.
The output weight of the minimum error was taken for final testing. Testing NRMSE error of 0.247 and MAPE error of 1.821 is obtained. The plot between final testing data and predicted data is shown on Figure 6.

6.2 One day prediction

Training phase

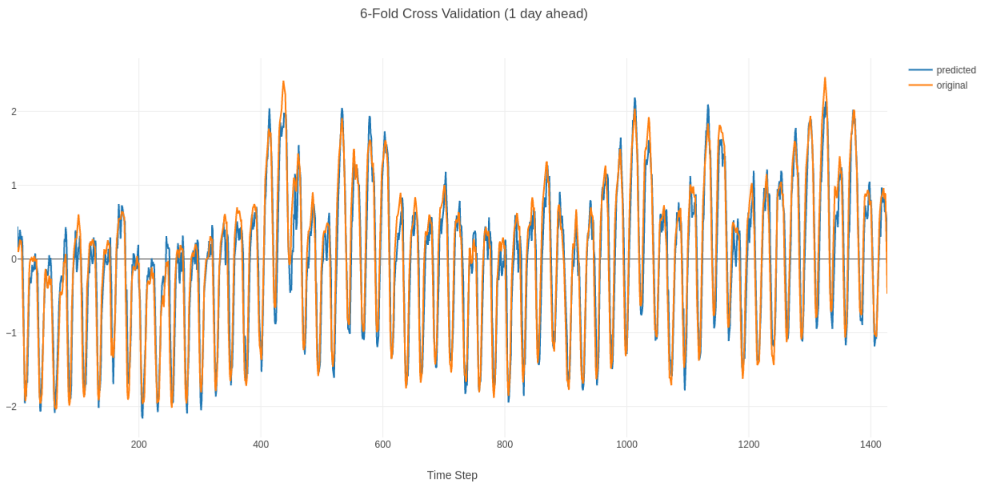

Figure 8: plot between training and validation in cross-validation.

Testing phase

After using the testing datasets, the desired test data and the output test data are plotted after calculation.

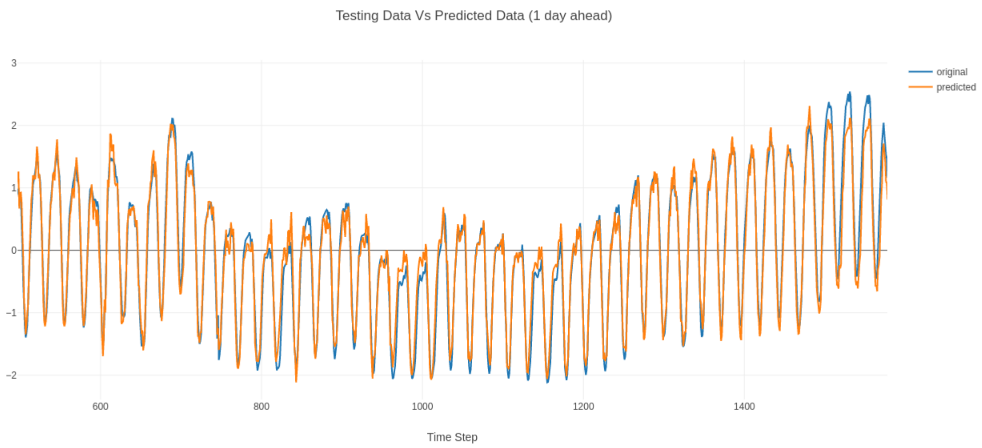

Figure 9: Result after testing the data.

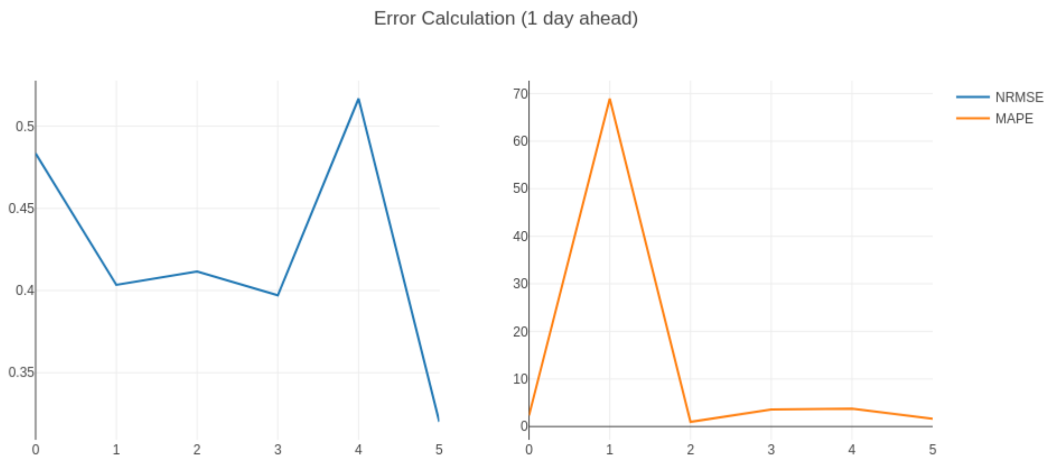

Figure 10: NRMSE of each fold.

The plot of each validation set and predicted result is shown in figure 8. The resulting error on each validation set is shown in Table 2.

	Partition 1	Partition 2	Partition 3	Partition 4	Partition 5	Partition 6
NRMSE	0.4835	0.403	0.4116	0.397	0.5167	0.320
MAPE	2.322	68.943	0.985	3.602	3.757	1.6395

Table 2: NRMSE and MAPE error in each partition for 1 day prediction.

The mean NRMSE error is 0.422 and the mean MAPE error is 13.54. The plot of NRMSE and MAPE of each partition are shown in figure 10.
The output weight of the minimum error was taken for final testing. Testing NRMSE error of 0.198 and MAPE error of 1.71 is obtained. The plot between final testing data and predicted data is shown on Figure 9..

6.3 One week prediction

Training phase

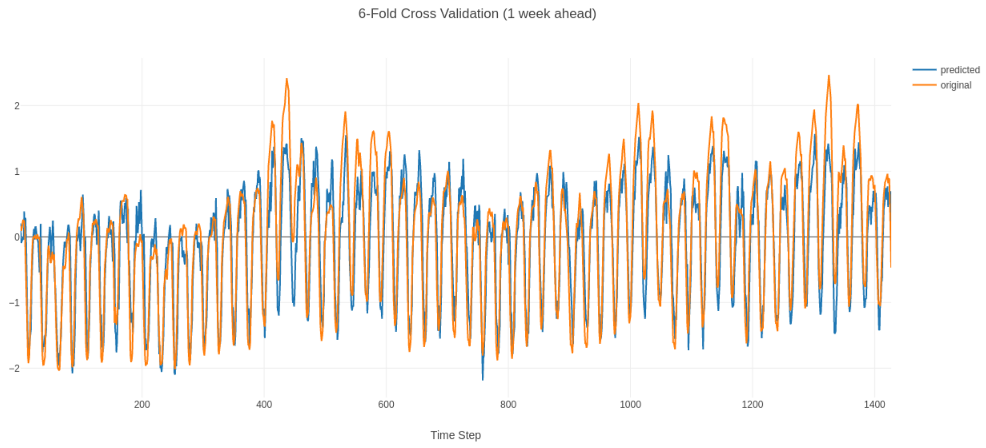

Figure 11: plot between training and validation in cross-validation.

Testing phase

After using the testing datasets, the desired test data and the output test data are plotted after calculation.

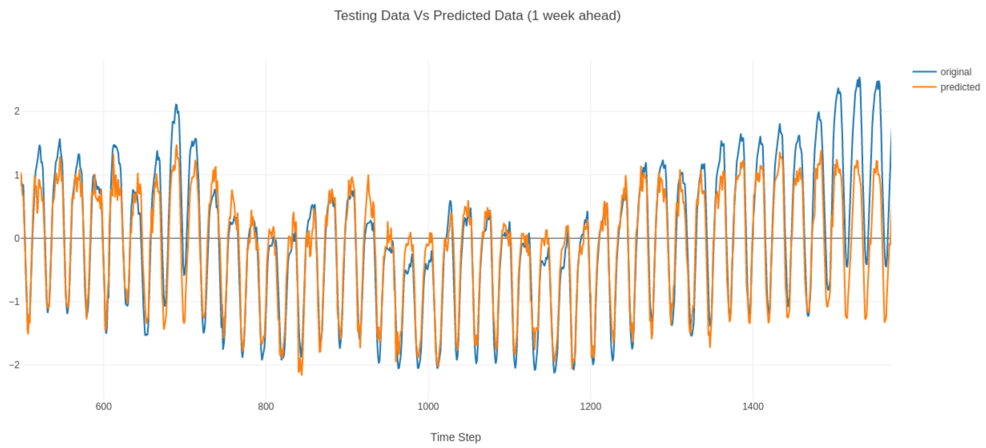

Figure 12: Result after testing the data.

Figure 13: NRMSE of each fold.

The plot of each validation set and predicted result is shown in figure 11. The resulting error on each validation set is shown in Table 3.

	Partition 1	Partition 2	Partition 3	Partition 4	Partition 5	Partition 6
NRMSE	0.6	0.49	0.509	0.531	0.69	0.428
MAPE	3.04	8.8	1.7	2.217	3.2	2.98

Table 3: NRMSE and MAPE error of 1 week data in each partition.

The mean NRMSE error is 0.545 and the mean MAPE error is 3.67. The plot of NRMSE and MAPE of each partition are shown in figure 13.
The output weight of the minimum error was taken for final testing. Testing NRMSE error of 0.34 and MAPE error of 1.91 is obtained. The plot between final testing data and predicted data is shown on Figure 12.

7 Discussion

After looking at the result of the prediction task, we can observe that the modal performs well with the prediction task of 1 hour more than 1 day and 1 week. As the time gap between the data increases the performance decreases which is quite normal. By applying cross-validation technique and manual optimization method, parameters were determined which helped to make prediction task better. Adding more features and extending the prediction task to different layers of data can be done such that the model performs well as a future task of the project.

A Data set Summary

In 4, a short glimpse of data set is given, where there are METER ID, TIME and KWH is provides with their values in its respective columns.

METER ID	TIME	KWH
NSLS_SPP	1:00	1113185.41
NSLS_SPP	2:00	1046074.53
NSLS_SPP	3:00	1002053.38
NSLS_SPP	4:00	1007000.65
NSLS_SPP	5:00	1008458.61

Table 4: Summary of a few data obtained from [9]

References

[1] Bo-Juen Chen, Ming-Wei Chang, and Chih-Jen lin. “Load forecasting using support vector Machines: a study on EUNITE competition 2001”. In: *IEEE Transactions on Power Systems* 19.4 (2004), pp. 1821–1830. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2004.835679.

[2] J. Y. Fan and J. D. McDonald. “A real-time implementation of short-term load forecasting for distribution power systems”. In: *IEEE Transactions on Power Systems* 9.2 (1994), pp. 988–994. ISSN: 0885-8950. DOI: 10.1109/59.317646.

[3] H. Jaeger. “Echo state network”. In: *Scholarpedia* 2.9 (2007). revision #183563, p. 2330. DOI: 10.4249/scholarpedia.2330. URL: http://www.scholarpedia.org/article/Echo_state_network.

[4] Jian-Kai, L., Cattani, C. and Cattani, C. “Power Load Prediction Based on Fractal Theory”. In: *Advances in Mathematical Physics* (2015). Article ID 827238, 2015. doi:10.1155/2015/827238, p. 6. URL: http://dx.doi.org/10.1155/2015/827238.

[5] S. J. Kiartzis et al. “A fuzzy expert system for peak load forecasting. Application to the Greek power system”. In: *2000 10th Mediterranean Electrotechnical Conference. Information Technology and Electrotechnology for the Mediterranean Countries. Proceedings. MeleCon 2000 (Cat. No.00CH37099)*. Vol. 3. 2000, 1097–1100 vol.3.

[6] Li, G.; Li, B.-J.; Yu, X.-G.; Cheng, C.-T. *The architecture of standard Echo State Network (ESN)*. [Online; accessed December 27, 2017]. 2015. URL: http://www.mdpi.com/energies/energies-08-12228/article_deploy/html/images/energies-08-12228-g001.png.

[7] Lukoŝeviĉius, M., and Jaeger, H. “Reservoir computing approaches to recurrent neural network training”. In: *Computer Science Review* 3.3 (2009), pp. 127 –149. ISSN: 1574-0137. DOI: https://doi.org/10.1016/j.cosrev.2009.03.005. URL: http://www.sciencedirect.com/science/article/pii/S1574013709000173.

[8] Maass W., Natschlaeger T., and Markram H. “Real-time computing without stable states: A new framework for neural computation based on perturbations.” In: *Neural Computation* 14(11) (2002). Online, pp. 2531–2560. URL: https://www.ncbi.nlm.nih.gov/pubmed/12433288.

[9] *Net System Load Shape*. Toronto Hydroelectricity data set, [Online; accessed December 27, 2017]. URL: https://www.torontohydro.com/SITES/ELECTRICSYSTEM/BUSINESS/YOURBILLOVERVIEW/NETSYSTEMLOADSHAPE/Pages/default.aspx.

[10] R.F. Engle, C. Mustafa, and J. Rice. “Modelling Peak Electricity Demand”. In: *Journal of Forecasting* (1992). 11:241–251.

[11] Wikipedia, the free encyclopedia. *Diagram of k-fold cross-validation with k=4.* [Online; accessed December 27, 2017]. 2016. URL: http://www.mdpi.com/energies/energies-08-12228/article_deploy/html/images/energies-08-12228-g001.pnghttps://en.wikipedia.org/wiki/Cross-validation_(statistics)#/media/File:K-fold_cross_validation_EN.jpg.

[12] Hong-Tzer Yang, Chao-Ming Huang, and Ching-Lien Huang. “Identification of ARMAX model for short term load forecasting: an evolutionary programming approach”. In: *Proceedings of Power Industry Computer Applications Conference*. 1995, pp. 325–330.