Improved Forecasting of Short Term Electricity Demand by using of Integrated Data Preparation and Input Selection Methods

Azadeh Arjmand[1,*], Reza Samizadeh[1] and Mohammad Dehghani Saryazdi[2]

[1] *Department of Industrial Engineering, College of Engineering, Alzahra University, Tehran, Iran.*
[2] *Department of Computer Engineering, College of Engineering, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran.*
**Corresponding author: a.arjmand@alzahra.ac.ir*

Manuscript received 01 April, 2018; Revised 18 September, 2018; accepted 15 November, 2018. Paper no. JEMT-1804-1077.

The main aim of this paper is to emphasize on the significant role of data pre-processing phase in improving the short-term load demand forecasting. Different transformation approaches including normalization, Zscore and Box-Cox methods are applied and various input selection methods including forward selection, backward selection, stepwise regression and principle component analysis are used to see how the combination of these pre-processing techniques will influence the performance of different parametric (ARIMA, ARIMAX, MLR) and non-parametric (NAR, NARX, SVR, ANFIS) predictors. The data has been collected from the daily load demand of Ottawa, Canada. It has been observed that the Box-Cox transformation significantly improved the performance of all predictors and the findings have demonstrated the superior role of exogenous variables in accuracy improvement of all predictors. In terms of MAPE, the value of 2.27% for ARIMA model improved to 1.75% with ARIMAX using temperature, and it decreased from 1.46% to 1.334% by means of NARX model using normalized PCA which is applied to normalized data. In an overall view, the non-parametric algorithms have had a considerable gain over parametric models and NARX network has the highest accuracy among all of the predictors.

Keywords: *Short-term load-demand forecasting, Pre-processing, Box-Cox transformation, NARX.*

http://dx.doi.org/10.22109/jemt.2018.126045.1077

1. Introduction

Electricity demand forecasting is known as one of the most important issues in energy management which plays an essential role in economic growth and development of countries. Since the operation of a wide range of industries and urban consumers mostly depend on the proper supply of electrical power, it is strictly required for every country to have an effective plan for the most reliable supplement with least costs. Load forecasting is divided into three main categories based on the time interval of the prediction: long term load forecasting (LTLF), medium term load forecasting (MTLF) and short-term load forecasting (STLF). In comparison with the first two categories, STLF is more considered in literature of load demand prediction due to its essential role in efficient daily planning and the operation cost reduction of power systems [1]. An accurate load forecast can conveniently carry the main power system operations such as maintenance scheduling, tariff rates adjustments and contract evaluations [2]. Moreover, it will improve the efficiency of the decisions made by energy managers and policy makers for having the most reliable energy system in future [3,4]. For STLF, the existing techniques are mainly categorized into two groups [5]. Parametric techniques such as time series techniques [4, 6, 7], linear regression [8], autoregressive moving average (ARMA) [9, 10] and stochastic time series [4, 11] and Non-parametric techniques such as artificial neural networks (ANNs) [12-17].

1.1. Related works

There are some review papers on the most commonly used techniques for STLF [18, 19]. Among forecasting models, statistical methods such as autoregressive moving average (ARMA) [9], time series techniques [6], linear regression [8] are known as strong predictors. However, if the behavior of the input data deviates from its normal condition (such as sudden accidents and deviations in exogenous variables), these methods aren't capable enough to quickly identify and support such abruptions. Nevertheless, they are still known as powerful forecasting tools. Vaghefi [20] proposed a model in which a multiple linear regression and a seasonal autoregressive moving average model are combined, so that it was possible to take advantage of two parametric methods simultaneously to forecast the short-term load demand. Kavousi-fard [21] established a hybrid model by combining ARIMA models with AI based algorithms. Bennette [12] developed a hybrid model of ARIMA and ANN based techniques which resulted in improvement in the forecasting of total energy consumption of the next day and determining demand.

For non-parametric forecasting studies, Lin [22] developed an ensemble model of Variational Mode Decomposition (VMD) and extreme learning machine (ELM) which were optimized by differential evolution (DE) algorithm. The results displayed a significant improvement in one-step and multi-step ahead forecasting of load demand. Zheng [23] established a hybrid model of similar day (SD) selection, empirical mode decomposition (EMD) and long short-term memory (LSTM) neural networks for prediction. The achievements revealed proper improvements in load forecasting. Buitrago [24] developed a hybrid model of open and closed-loop form of non-linear autoregressive neural network with exogenous variables (NARX) models. In terms of average error, the proposed model achieved an improvement of 30% in comparison with the feed-forward ANNs and ARMAX model.

The idea of support vector machine (SVM) was first created by Vapnik in 1996 [25]. Among related works, Pellegrini [26] proposed a SVR model for the nonlinear dynamic behavior of customer load demand without any assumption for the stationary nature of the input data. Chen et al. [27] used the SVR to forecast the load demand in which the use of temperature as input variable could significantly improve the accuracy.

Adaptive Neuro-Fuzzy interface System (ANFIS) method was first introduced by Jang [28] who took advantage of both ANNs and FL systems to establish a strong prediction tool with minimum error. Among ANFIS papers, Yang et al. [29] proposed a hybrid model based on ANFIS and an improved neural network algorithm which could deal with linearity, nonlinearity and seasonality problems in STLF. Chevik and Chunkas [30] have developed an ANFIS model to forecast hourly load demand of Turkey in a one-year horizon and the historical load and temperature have been used as input data.

The capability and efficiency of both categories in forecasting are eminent [31]. However, recent studies reveal that the artificial intelligence algorithms have shown more eminent performance in forecasting [32], especially in cases that the normal conditions are affected by sudden abruptions (human impacts, social events and meteorological changes) [33]. Yet in some cases, the parametric models, such as ARIMA, have represented an impressive performance in predicting the load consumption because of their dynamic structure [34].

In addition to developing the most efficient predictors, two aspects are also of importance in establishing an accurate prediction: Considering all of the factors which are effective on load demand variations (endogenous and exogenous) and improving the quality of the input data by using appropriate pre-processing methods. Input selection can significantly improve the prediction accuracy. Among exogenous variables, Weather factors (temperature, humidity, wind speed) and historical data are mostly considered in load demand forecasting [12, 35]. In addition to well-known input selection techniques, in some cases, this procedure is done based on trial and error [36, 37]. The effective factors are those in which a significant correlation between their values and the load consumption is investigated. Bennett [12] used SR method for selecting the variables for STLF. Massana [38] applied a heuristic method in which the whole space of features is searched and the redundant variables are removed. Zheng [23] applied an Xgboost algorithm to evaluate the importance of exogenous features and selected temperature and next-day pick load as the most effective variables.

In addition to Input selection, the quality of the input data also plays an essential role in achieving a precise prediction model and has a remarkable effect on the performance of the predictors. Pre-processing techniques not only improve the accuracy of the prediction, but also are appropriate for the characteristics of the experimental model [39]. In time series prediction, covariance stationary assumption guaranties that the mean and covariance of the process is finite and time invariant. The non-stationary characteristics of load demand series can be removed by pre-processing step which can significantly improve the quality of input data. Although data-preprocessing methods are strictly considered in some literature, but, in some cases, using covariance stationary methods are ignored [34, 20, 40, 41] or the regular normalization technique with discrete uniform distribution is applied [42-44]. However, there are other powerful transformation methods which should be taken into account.

This paper first investigates the significant effects of different data pre-processing methods on the performance of algorithms from both parametric and non-parametric categories. Second, the effects of a group of input selection techniques on the performance of the predictors are considered to see how various selection techniques influence the performance of algorithms and the accuracy of the prediction. The aim is to compare the performance of various algorithms with different data pre-processing and input selection methods and confirm that this simple but essential step cannot be ignored in forecasting problems. In summary, the contribution of the paper is given as below:

- Applying various data pre-processing methods to develop precise STLF models.
- Using different input selection methods to involve the most effective factors in load demand prediction.
- Investigating the effects of different data pre-processing and input selection techniques on both parametric and non-parametric predictors.
- Considering the mutual effects of data pre-processing and input selection methods to identify the best combination for forecasting.

The rest of the paper is organized as follows: the theory of the parametric and non-parametric algorithms is brought in section 2. In section 3, the data pre-processing techniques are discussed. In section 4, the input selection approaches are introduced. Next, design of the experiment is given in section 5. The prediction results are brought in section 6, and the related discussions are presented in section 7. A brief conclusion is provided in section 8 and finally some future directions are given in section 8.

2. Applied statistical and artificial methods

In this paper, some of the parametric algorithms are selected from a group of ARMA-based models including ARIMA and ARIMAX and the multivariate linear regression (MLR). The non-parametric algorithms are chosen from MLP-based methods including Support Vector Regression (SVR), Nonlinear AutoRegressive model with Exogenous variables (NARX) and Nonlinear AutoRegressive model (NAR). The NAR networks are designed to see how the performance of the model will change without the exogenous variables.

2.1. Adaptive neuro-fuzzy interface system (ANFIS)

In addition to the abovementioned non-parametric algorithms, an Adaptive Neuro-Fuzzy Interface System (ANFIS) model is developed to take advantage of fuzzy logic and ANN combination and see how it works with various scenarios of input selection and data pre-processing.

Suppose that there are two inputs x_1 and x_2 and one output Y. Based on the first-order Sugeno fuzzy model, the common fuzzy rules with two if-then expressions are determined as given below:

R_1: If x_1 is α_1 and x_2 is β_1, then $y_1=p_1x_1+q_1x_2+r_1$

R_2: If x_1 is α_2 and x_2 is β_2, then $y_2=p_2x_1+q_2x_2+r_2$

where α_i and β_i (i=1,2) are the fuzzy sets with membership functions $\mu\alpha_i$ and $\mu\beta_i$. Figure 1 displays the structure of the equivalent ANFIS function with two inputs. Each layer proceeds with the following evaluations:

A. Rule premises evaluation

$$w_i = \mu\alpha_i(x_1)\mu\beta_i(x_2) \qquad i=1,2 \tag{1}$$

B. Implication evaluation and final output

$$Y(x_1,x_2)=\frac{w_1y_1+w_2y_2}{w_1+w_2} \tag{2}$$

which can be rewritten as:

$$Y(x_1,x_2)=\overline{w_1}y_1+\overline{w_2}y_2 \tag{3}$$

where

$$\overline{w_i}=\frac{w_i}{w_1+w_2} \tag{4}$$

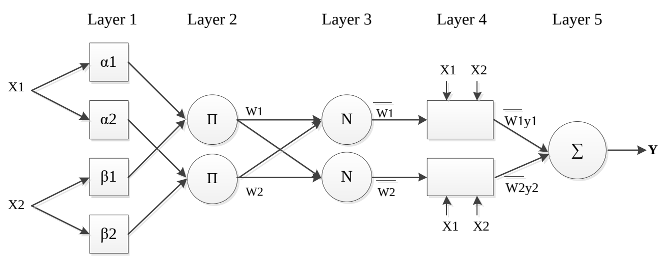

Fig 1: ANFIS structure with two inputs and one output

3. Data pre-processing

In this study, the dataset of daily electricity demand from Ottawa, Canada is used and the records are gathered from January 1, 2013 to January 7, 2016, as illustrated in Figure 2.

The histogram and the statistics of the actual load demand are given in Figure 3 and Table 1 which show a positive skewness (far different from zero) and kurtosis (higher than 3). It indicates an asymmetric distribution for the raw data and thus, the actual data is not normally distributed and requires to be normalized.

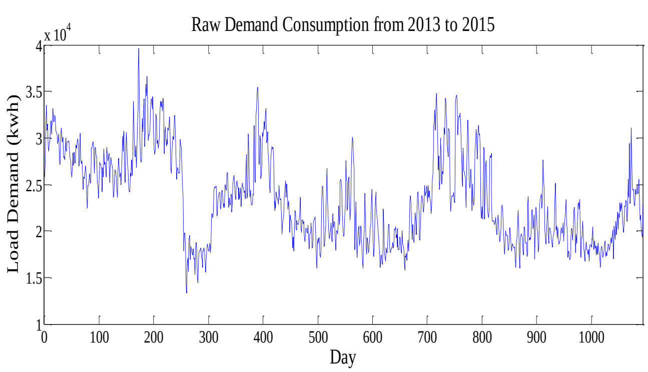

Fig. 2. The actual load in Ottawa from Jan. 1, 2013 to Jan. 7, 2013

In addition to historical demand, six exogenous variables including average daily temperature (T), humidity (H), dew point (DP), visibility (V), sea level pressure (SLP) and wind speed (WS) are gathered to use for multivariate forecasting methods. Fortunately, the dataset is complete and no information is missed. The outlier issues are supposed to be ignored, that is, the odd data points (sudden increase/decrease) are kept in the dataset to evaluate the performance of the predictors in facing such abruptions.

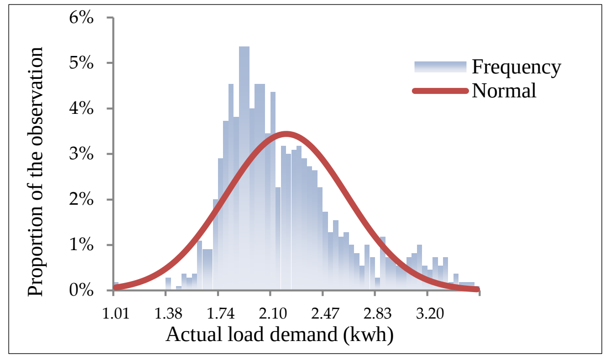

Fig. 3. Histogram of actual load demand

Table 1. Summary statistics of the actual load demand dataset

Parameter	Value
Average	21974.82
Median	20997.5
Maximum	35432
Minimum	9934
Std. Dev.	4222.287
Skewness	0.93
Kurtosis	3.51
1st Quartile	18932.5
3rd Quartile	24033.75

3.1. Stationary process and data transformation

Before developing the predictor models, data pre-processing step is essential to improve the precision of the prediction. A covariance stationary time series is one of the main initial assumptions for applying ARIMA models. A covariance stationary process in Box-Jenkins model [45] is defined as a series in which the mean and variance do not change over the time. For a non-stationary process, a number of techniques are suggested to stabilize the mean and variance such as differencing and Box-Cox transformation [46]. In this study, various data normalization methods including Max-Min normalization, Zscore, first difference and Box-Cox are applied to stabilize the mean and variance of the process and evaluate the performance of predictors with various transformed data. In the following, these methods are briefly explained.

- *Min-Max normalization*

Assume that the transformed data is going to be in the interval [$x'_{\min}, x'_{\max}$]. Then the normalized data is achieved as displayed in Eq. (5),

$$x'_{new}=\left(\frac{x_{old}-x_{\min}}{x_{\max}-x_{\min}}\right)(x'_{\max}-x'_{\min})+x'_{\min} \tag{5}$$

- *Zscore normalization*

In this method given in Eq. (6), the mean and standard deviation of the transformed data are zero and one, respectively:

$$x_{new}=\frac{x_{old}-mean_{old}}{std_{old}} \tag{6}$$

- *First difference method*

In Box-Jenkins model, differencing is used to obtain a stationary process. Differencing can be applied in various orders to obtain a stable process. The first difference is given in Eq. (7).

$$x'_t=x_t-x_{t-1} \tag{7}$$

- *Box-Cox transformation*

In addition to mean, the variance stationary time series is required for time series prediction. In this regard, Box-Cox transformation, given in (8), is suggested to stabilize the variance of the process [40]:

$$x_t'(\lambda) = \begin{cases} \dfrac{(x_t)^{\lambda}-1}{\lambda}, & if \;\; \lambda \neq 0 \\ \log(x_t), & if \;\; \lambda = 0 \end{cases} \quad \textbf{(8)}$$

The values of 0, 0.5 and $\frac{1}{3}$ are widely used which the last two transformed values are called as square and cubic roots.

4. Input selection

For multivariate predictions, input selection can significantly improve the performance of predictors. Identification of the most influencing factors facilitates the data gathering, especially for long period predictions which helps to collect fewer data. Among different techniques for data reduction, in this research, forward selection (FS), backward selection (BS), stepwise regression (SR) and principal component analysis (PCA) are the methods applied for input selection and their performance are compared to see how they work with parametric and non-parametric predictors.

FS, BS and SR are regression-based models which consider the correlation between the input and dependent variables. PCA is known as a technique for reducing/removing inefficient variables from the original dataset [47] (Azadeh and Ebrahimipour, 2004). PCA investigates for a new set of variables (principal components) which are defined as an uncorrelated linear combination of the original input variables. It is expected that the performance of PCA improves when the initial exogenous variables have least variance. In this case, the PCA is applied on exogenous variables in two forms: Raw and normalized values of exogenous input data. In summary, Figure 4 illustrates the overall methodology of the paper.

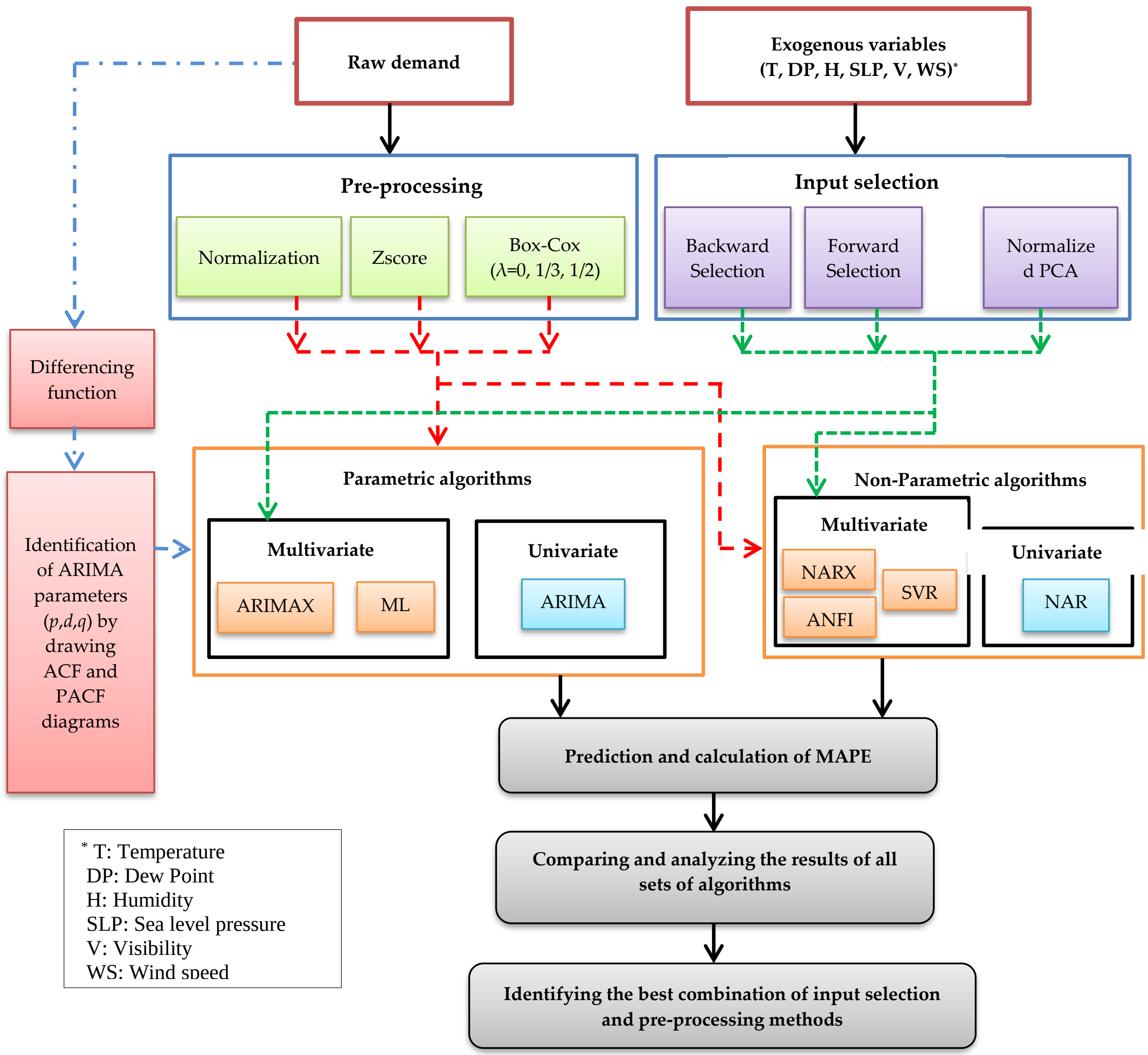

Fig. 4. Graphical abstract of the overall methodology

5. Design of experiment

To design the experiment, the daily historical load demand and average weather data are collected. The total number of instances is 1102 covering the data from January 1, 2013 to January 7, 2016. The original load demand data are obtained from the Independent Electricity System Operator (IESO) website of the Ontario's power system (http://www.ieso.ca). All prediction models are tested for a 7-day-ahead period.

5.1. Pre-processing

ACF and PACF plots are useful tools indicating whether the process is covariance stationary or not. Figure 5 displays the ACF and PACF of raw data.

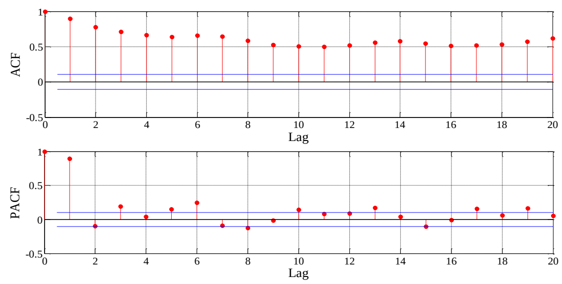

Fig. 5. ACF and PACF of raw data before pre-processing

If ACF plot decays slowly and PACF displays a sudden cut off, it is a sign of non-stationary process. To cope with it, differencing is a well-known method to alleviate the variations in the process and make it stationary. Figure 6 displays the series of differenced raw data and Figure 7 shows how differencing effects on the ACF and PACF plots.

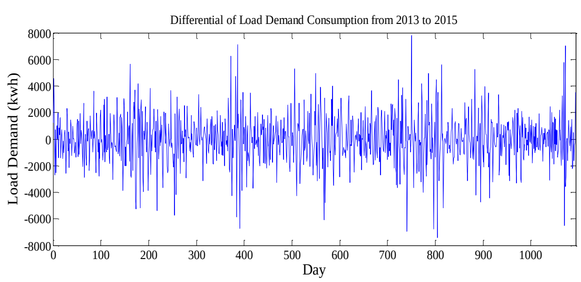

Fig. 6. The trend of differenced load demand in Ottawa from Jan. 1, 2013 to Jan. 7, 2016

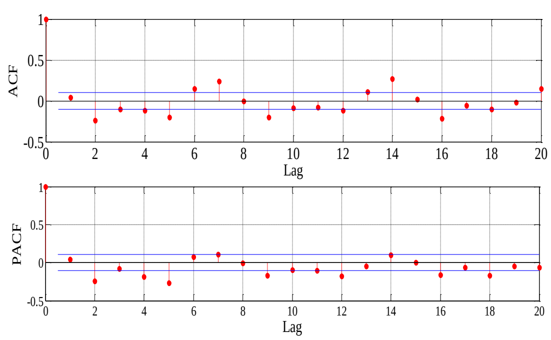

Fig. 7. ACF and PACF of differenced data

5.2. Input selection

In prediction, it is common to select those variables which have the highest correlation with the response variable. Figure 8 illustrates the correlation of exogenous weather variables with the daily load demand.

Based on the given correlation coefficients, T and DP have the highest correlation with load demand. The rest of the input selection methods have selected various combinations; T and H are selected by FS and SR methods and H and DP are chosen by BS technique. For PCA approach, three PCAs are considered which are able to cover at least 88.7% of the variance between exogenous variables. The results further indicate how these various techniques will behave with different algorithms.

5.3. Parameters Setting for SVR

SVR uses kernel functions to transform the data into a new feature space and then performs a linear regression as given in (9):

$$f(x) = \sum_{i=1}^{n} (\alpha_i^* - \alpha_i) k(x_i, x) + b \tag{9}$$

where α_i and α_i^* are the Lagrangian multipliers, $k(x_i, x)$ is the kernel function.

There are several kernel functions with various characteristics such as radial basis, linear, polykernel and Pearson VII universal kernel (PUK) functions. Among all, PUK is the most suitable function for mapping and generalizing the data points with the form given in (10):

$$K(x_i, x_j) = \frac{1}{[1 + (2\sqrt{\|x_i - x_j\|^2}\sqrt{2^{(\frac{1}{\omega})-1}} / \sigma)^2]^{\omega}} \tag{10}$$

The related parameters of ω and σ in PUK and parameter C for SVR should be specified. Parameter C plays as a controller which penalizes the mis-classified cases. Although there's not a proved theory for C specifications, but one reasonable idea is to set it around the range of output values [38].

In order to set the parameters of the SVR model, grid search (GS) is used as a simple and fast method which is appropriate for the size of the problem of this paper. Grid search explores among pairs of parameters until it finds the best. By using of GS, the values of 0.4 and 14 are specified as ω and σ. Since the output data is pre-processed and transformed with different methods of normalization, its value varies in different cases as given in Table (2):

Table 2. The parameters estimated for SVR model

	Raw data	Normalization	Zscore	Box-Cox		
				L=0	L=1/3	L=1/2
Parameter *C*	19000	0.3	1	4	82	300

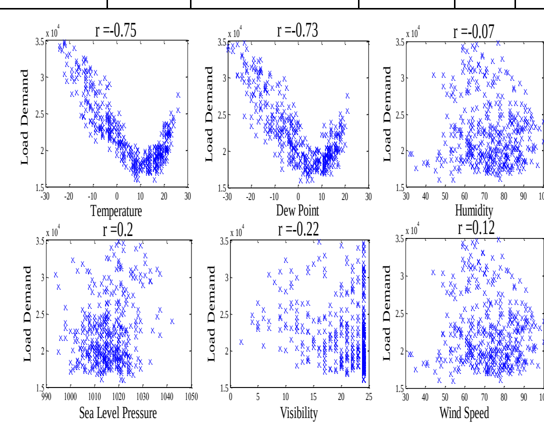

Fig. 8. The correlation scatter plot for six exogenous variables

6. Prediction results

The performance quality is quantified by means of mean absolute percentage error (MAPE) metric to be compared with the conventional studies. The MAPE is calculated as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|A_i - F_i\right|}{A_i} \tag{11}$$

where A_i and F_i stand for actual and forecast value, respectively.

6.1. Prediction without Exogenous Variables

Table 3 and Table 4 present the performance of ARIMA (1,1,1) and NAR network in terms of MAPE percentage. Generally, NAR has shown a better performance over ARIMA. For both predictors, the Box-Cox is the best transformation method in which the lower value for parameter λ leads to a better precision. Zscore technique, on the other hand, had displayed the weakest performance in comparison with the rest of the pre-processing approaches.

Table 3. The value of MAPE(%) for ARIMA (1,1,1)

	Data Transformation Method					
	Raw Data	Normalization [0,1]	Zscore	Box-Cox Transformation		
				λ=0	λ=0.33	λ=0.5
ARIMA (1,1,1)	11.50	27.56	64.78	2.27	5.35	7.23

Table 4. The value of MAPE(%) for NAR network

	Data Transformation Method					
	Raw Data	Normalization [0,1]	Zscore	Box-Cox Transformation		
				λ=0	λ=0.33	λ=0.5
NAR	5.98	11.71	29.59	1.46	2.53	3.12

6.2. Prediction with Exogenous Variables

The results of the predictions which consider the exogenous variables for parametric and non-parametric algorithms are given in Tables 5 to 9. The results reveal that those exogenous variables with higher correlation coefficient are more capable of improving the prediction accuracy rather than using all variables. In addition, the variables selected by FS, BS, SR and PCAs are not the best ones in comparison with T and DP.

Table 5. The value of MAPE(%) for ARIMAX (1,1,1)

		Data Transformation Method					
		Raw Data	Normalization [0,1]	Zscore	Box-Cox Transformation		
					λ=0	λ= 0.33	λ= 0.5
Exogenous Variable(s)	All	7.58	17.05	36.53	2.08	3.88	5.10
	T	7.73	14.00	28.68	1.75	3.48	4.50
	T,DP	6.41	13.92	28.57	2.56	6.26	8.51
	T,H*	14.89	14.18	29.42	2.61	6.41	8.70
	H,DP**	14.80	13.94	28.85	2.61	6.43	8.73
	PCA	7.66	15.83	33.46	1.83	3.75	4.89
	PCA***	7.62	15.74	33.25	1.82	3.73	4.87

* Forward Selection, Stepwise Regression
** Backward Selection
*** Normalized PCA

Unlike ARIMAX model, the results given in Table 6 reveal that the MLR requires all exogenous variables to improve its performance.

Table 6. The value of MAPE(%) for MLR

		Data Transformation Method					
		Raw Data	Normalized [0,1]	Zscore	Box-Cox Transformation		
					λ=0	λ= 0.33	λ= 0.5
Exogenous Variable(s)	All	5.94	15.05	43.77	1.46	2.77	3.68
	T	6.70	16.75	47.94	1.57	3.08	4.12
	T,DP	6.74	16.84	48.21	1.57	3.08	4.12
	T,H*	6.75	16.87	48.36	1.57	3.09	4.13
	H,DP**	6.78	16.77	47.47	1.59	3.17	4.21
	PCA	6.80	17.01	47.41	1.56	3.11	4.16
	PCA***	6.79	16.95	47.16	1.56	3.10	4.15

* Forward Selection, Stepwise Regression
** Backward Selection
*** Normalized PCA

Table 7 gives the final results for SVR model in which Temperature as input variable and Box-Cox method as normalization data resulted in the best prediction accuracy. In Table 8, the achievements reveal that the ANFIS model has the best performance with Temperature and Dew point as the variables with highest correlation with load data.

Table 7. The value of MAPE (%) for SVR

		Data Transformation Method					
		Raw Data	Normalized [0,1]	Zscore	Box-Cox Transformation		
					λ=0	λ= 0.33	λ= 0.5
Exogenous Variable(s)	All	5.81	15.02	26.53	1.84	3.05	4.02
	T	4.72	13.08	15.34	1.44	2.53	3.59
	T,DP	5.17	13.94	16.25	1.47	2.59	3.61
	T,H*	5.33	14.27	16.93	1.49	2.61	3.62
	H,DP**	5.39	14.36	17.52	1.52	2.63	3.65
	PCA	5.64	14.63	21.64	1.75	2.84	3.77
	PCA***	5.56	14.42	18.71	1.61	2.75	3.72

* Forward Selection, Stepwise Regression
** Backward Selection
*** Normalized PCA

Table 8. The value of MAPE (%) for ANFIS

		Data Transformation Method					
		Raw Data	Normalized [0,1]	Zscore	Box-Cox Transformation		
					λ=0	λ= 0.33	λ= 0.5
Exogenous Variable(s)	All	5.35	11.02	24.49	1.51	3.45	3.81
	T	4.23	8.94	11.98	1.46	2.39	3.44
	T,DP	4.03	8.61	11.62	1.41	2.16	3.26
	T,H*	4.52	10.44	23.53	1.49	2.67	3.55
	H,DP**	4.31	9.27	18.61	1.48	2.54	3.46
	PCA	5.41	11.74	24.87	1.53	3.76	3.97
	PCA***	5.16	10.51	24.12	1.50	3.23	3.64

* Forward Selection, Stepwise Regression
** Backward Selection
*** Normalized PCA

Finally, the performance of NARX, as given in Table 9, displays a high superiority over the rest of the algorithms. The combination of Normalized PCAs as input variables and Box-Cox transformation of raw data have resulted in the best prediction with highest accuracy.

Figure 9 illustrates the performance of the abovementioned NARX model in some period (250 days) in the dataset.

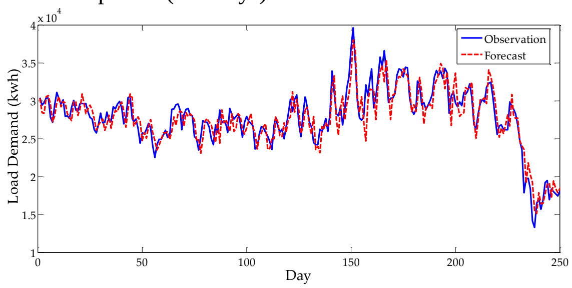

Fig 9. NARX network performance with Box-Cox Transformation and Normalized PCA

Table 9. The value of MAPE(%) for NARX network

		Data Transformation Method					
		Raw Data	Normalized [0,1]	Zscore	Box-Cox Transformation		
					λ=0	λ=0.33	λ=0.5
Exogenous Variable(s)	All	3.71	7.56	22.09	1.36	1.86	2.50
	T	4.22	9.46	20.88	1.35	2.11	2.71
	T,DP	4.98	10.37	10.56	1.39	2.26	2.71
	T,H*	3.87	8.06	22.20	1.338	1.86	2.70
	H,DP**	4.69	10.18	15.95	1.41	2.36	3.09
	PCA	4.23	9.14	22.58	1.38	1.61	2.67
	PCA***	3.49	7.44	9.95	1.334	1.46	2.43

* Forward Selection, Stepwise Regression
** Backward Selection
*** Normalized PCA

7. Discussion

A comparison of the findings presented in Tables 5 to 9 shows the significant role of applying Box-Cox as a transformation method on accuracy improvement in daily load demand prediction especially for λ=0. The nonlinear transform of the raw data leads to a covariance stationary process in which the minimum value of λ results in more precision. Figure 10 illustrates how the Box-Cox transformation behaves with positive inputs and different values of parameter λ (Lambda). The Box-Cox function displays an almost linear behavior with lower values of Lambda and it is preferred to predict values with least variation.

In a different point of view, the achievements in Table 10 reveal how the Box-Cox method could better make the distribution of the dataset close to normal. Although the normalization approach has significantly reduced the kurtosis close to zero, but the skewness is still too high. The histograms of the transformed data by Box-Cox and normalization methods are displayed respectively in Figure 11 and Figure 12.

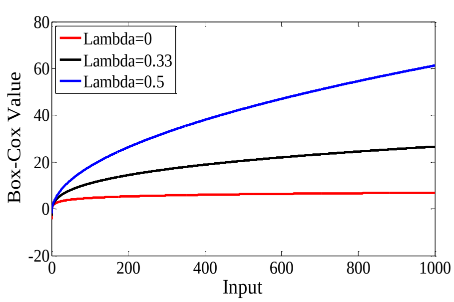

Fig 10. The Box-Cox transformation function with various values of Lambda

Table 10. Summary statistics of the actual and transferred load demand dataset

Parameter	Raw data	Normalized	Zscore	Box-Cox		
				λ=0	λ=0.33	λ=0.5
Average	21974	-0.2344	3E-16	4.33	80.70	293.18
Median	20997	-0.2797	-0.24	4.32	79.76	287.81
Maximum	35432	1	3.17	4.54	95.53	374.46
Minimum	9934	-1	-2.63	3.99	61.49	197.33
Std. Dev.	4222	0.4380	0.99	0.07	5.18	27.65
Skewness	0.93	0.72	0.96	0.43	0.61	0.69
Excess Kurtosis	0.51	0.04	0.51	0.23	0.21	0.25
1st Quartile	18932	-0.5672	-0.72	4.27	76.95	273.19
3rd Quartile	24033	0.0011	0.46	4.38	83.57	308.05

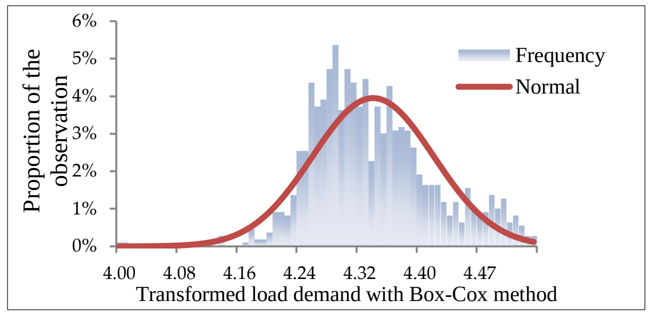

Fig 11. Histogram of transformed load demand by Box-Cox method

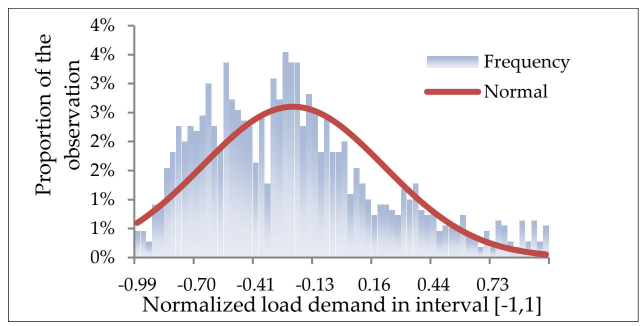

Fig 12. Histogram of transformed load demand by normalization method

Based on the findings in Tables 5 to 9, the performance of all predictors has improved with normalized PCA in comparison with the PCA on raw data. When the data is normalized, the variance within variables decreases and the components cover a higher portion of total variance. However, PCA has not been successful in improving the accuracy of ARIMAX and MLR. The reason may be the linear nature of these regression methods which are not compatible with the principal components as inputs. Since the MLR only uses exogenous variables, its performance enhanced with all input variables. On the contrary, ARIMAX performed accurate with one or two variables which have high correlation with the response variable.

The performance of SVR displayed an improvement with T as the variable with highest correlation with load demand and the PCA approach has not been successful in decreasing the prediction accuracy. The function of SVR model requires transforming the data into a new feature space, in this case, the components of PCA which are already the transformed form of original input data may not be capable to be retransformed and used as a representative regression variable and temperature which could better improve the accuracy of the model.

Similarly, PCA approach has not been appropriate for ANFIS which best improved with T and DP. Although its performance has been better than SVR and NAR, but its prediction accuracy has not been accurate as NARX. In comparison with NARX performance in Figure 9, the performance of ANFIS model for a period of 250 days, given in Figure 13, is less accurate than NARX network. The reason may be that the fuzzy variables and membership functions are not capable tools for modeling the complex relationship existing between the input data and load demand pattern.

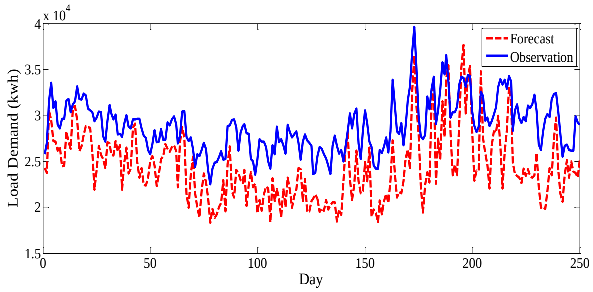

Fig 13. ANFIS performance with Box-Cox Transformation, T and DP as input variables

In general, two aspects are of importance in improving the performance of the predictors. First, the part of algorithms which uses the historical values of load demand and second, the part in which the exogenous variables are included. The first part requires a suitable method of transformation which makes the process covariance stationary, such as Box-Cox method which displayed a great improvement in prediction accuracy. For the second part, the input selection depends on the algorithm. For parametric algorithms, it is better to pay more attention to the correlation of the inputs with response variable and mostly apply such variables with their original value. For non-parametric algorithms such as neural networks, the ability of learning the behavior of the response variable enables them to take advantage of all the information hidden in the inputs. In the other words, such algorithms are capable of identifying very complex relationships existing among the input and response variable. As a result, principal components which carry a piece of information can be more capable in improving the performance of intelligent prediction algorithms.

The results show that the application of exogenous variables has led to more accurate prediction, but this is not necessarily always true. For instance, in comparison with the ARIMA, the accuracy of ARIMAX has improved in all conditions except for some few cases that the variables are selected by FS and BS. Comparing NARX and NAR results, the performance of NARX is far better than NAR, especially for the Zscore which has improved its MAPE from 29.59 to 9.95. The reason for such improvement may be the nature of AI algorithms in which the learning process improves when the inputs carry a set of useful information about the output variable.

Superior performance of MLR over ARIMAX and ARIMA reveals that the exogenous variables can add a great deal of helpful information to the process of prediction. It should be noted that in dealing with parametric and regression methods, the input variables should be wisely selected. In such models, the correlation of variables plays the most important role in input selection and redundant inputs can negatively affect the accuracy of the prediction.

The superiority of artificial intelligence algorithms over the parametric and regression models has been proved in many conventional studies of load demand prediction. The achievements of this paper again show that AI algorithms performed much better than ARIMA models in all cases (with/without exogenous variables). The reason for this superiority may be the complex nature of such time series which requires a powerful tool for recognition and learning of the sophisticated behavior of the input and output variables.

Finally, comparing the performance of transformation methods for all algorithms, the capability of Box-Cox transformation is prominent in all cases. Among the rest, Zscore displays the weakest performance in accuracy improvement. The reason may be the complex stochastic nature of the load demand which is not suitable for being transformed in a simple normal distribution form. In addition, though the normalization is better than the Zscore method, but it has not been so successful in making improvements in prediction accuracy.

8. Conclusion

The aim of this paper is to emphasize on the significant role of data pre-processing step in STLF. Five transformation methods are applied on the response variable and various input selection approaches including FS, BS, SR and PCA have been used to see how the combination of such pre-processing techniques would effect on the performance of the predictors. seven different kinds of predictors including parametric (ARIMA, ARIMAX and MLR) and non-parametric (NAR, SVR, ANFIS and NARX) models have been used to evaluate their performance in various conditions.

The experiment has been tested on dataset of daily load demand of Ottawa in Canada and the experimental results revealed that the Box-Cox transformation method extremely improved the accuracy of the prediction than normalization and Zscore methods which are widely used in conventional studies. The lower value of parameter λ also leads to a more constant, covariance stationary process and accurate predictions.

In addition, the performance of parametric algorithms has been improved with those exogenous variables which are highly correlated with the load demand.. The superior performance of MLR over ARIMA model proved that the exogenous variables also carry important information for predicting the behavior of the load demand. The findings showed that the PCA will result in better achievements if it is applied on normalized data than the raw ones and NARX model could outperform all the rest. Additionally, the performance of non-parametric algorithms has been far better than the parametric ones in load demand prediction which is supposed to be among the complex problems of time series prediction.

To summarize the achievements of this paper, the pre-processing, data transformation and input selection steps for time series prediction are the simplest but very essential activities which should be highly considered in analyzing sophisticated time series such as daily load demands. Table 11 gives a brief comparison between some recent conventional studies with the current paper and proves how the achievements of this paper are superior to some of previous studies.

Table 11. Comparison results of MAPE in some recent conventional researches

Paper	Algorithm		Exogenous Variables	Input selection	pre-processing	MAPE (%)
	parametric	non-parametric				
[37]	*	*	*		Normalization	3.08
[25]	*		*	*	Differencing	2.91
		*			Normalization	5.01
[12]	*		*	*	-	4.21
		*	*	*	Normalization	3.87
[29]		*	*	-	Normalization	1.49
Current paper	*		*	*	Box-Cox	1.46
		*	*	*	Box-Cox	1.334

9. Future directions

For future directions, the effects of further pre-processing techniques such as wavelet techniques and heuristic algorithm on various prediction algorithms can be investigated. Moreover, using various datasets from different geographical zones may reveal how weather variables have influence on the quality of forecasting of different kinds of predictors.

References

[1] Falomir, Z., and A. Olteţeanu. "Logics based on qualitative descriptors for scene understanding." *Neurocomputing* 161 (2015): 3-16.

[2] Methaprayoon, K., W. J. Lee, P. Didsayabutra, J. Liao, and R. Ross. "Neural network-based short term load forecasting for unit commitment scheduling." In *IEEE Technical Conference on Industrial and Commercial Power Systems, 2003.*, pp. 138-143. IEEE, 2003.

[3] Ruzic, S., A. Vuckovic, and N. Nikolic. "Weather sensitive method for short term load forecasting in electric power utility of Serbia." *IEEE Transactions on Power Systems*18, no. 4 (2003): 1581-1586.

[4] Kamel, N., and Z. Baharudin. "Short term load forecast using Burg autoregressive technique." In *2007 International Conference on Intelligent and Advanced Systems*, pp. 912-916. IEEE, 2007.

[5] Raza, M. Q., Z. Baharudin, B. Islam, M. A. Zakariya, and M. M. Khir. "Neural network based stlf model to study the seasonal impact of weather and exogenous variables." *Research Journal of Applied Sciences, Engineering and Technology* 6, no. 20 (2013): 3729-3735.

[6] Marín, F. J., and F. Sandoval. "Short-term peak load forecasting: Statistical methods versus artificial neural networks." In *International Work-Conference on Artificial Neural Networks*, pp. 1334-1343. Springer, Berlin, Heidelberg, 1997.

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

[8] Amral, N., C. S. Ozveren, and D. King. "Short term load forecasting using multiple linear regression." In *2007 42nd International universities power engineering conference*, pp. 1192-1198. IEEE, 2007.

[9] Chen, J., W. Wang, and C. Huang. "Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short-term load forecasting." *Electric Power Systems Research* 34, no. 3 (1995): 187-196.

[10] Barak, S., and S. S. Sadegh. "Forecasting energy consumption using ensemble ARIMA–ANFIS hybrid algorithm." *International Journal of Electrical Power & Energy Systems* 82 (2016): 92-104.

[11] Chakhchoukh, Y., P. Panciatici, and L. Mili. "Electric load forecasting based on statistical robust methods." *IEEE Transactions on Power Systems* 26, no. 3 (2011): 982-991.

[12] Bennett, C., R. Stewart, and J. Lu. "Autoregressive with exogenous variables and neural network short-term load forecast models for residential low voltage distribution networks." *Energies* 7, no. 5 (2014): 2938-2960.

[13] Hernandez, L., C. Baladrón, J. Aguiar, B. Carro, A. Sanchez-Esguevillas, and J. Lloret. "Short-term load forecasting for microgrids based on artificial neural networks." *Energies* 6, no. 3 (2013): 1385-1408.

[14] Amjady, N., and F. Keynia. "A new neural network approach to short term load forecasting of electrical power systems." *Energies* 4, no. 3 (2011): 488-503.

[15] Azadeh, A., S. Davarzani, A. Arjmand, and M. Khakestani. "Improved prediction of household expenditure by living standard measures via a unique neural network: the case of Iran." *International Journal of Productivity and Quality Management* 17, no. 2 (2016): 142-182.

[16] Chaturvedi, D. K., A. P. Sinha, and O. P. Malik. "Short term load forecast using fuzzy logic and wavelet transform integrated generalized neural network." *International Journal of Electrical Power & Energy Systems* 67 (2015): 230-237.

[17] Lou, C. W., and M. C. Dong. "A novel random fuzzy neural networks for tackling uncertainties of electric load forecasting." *International Journal of Electrical Power & Energy Systems* 73 (2015): 34-44.

[18] Tzafestas, S., and E. Tzafestas. "Computational intelligence techniques for short-term electric load forecasting." *Journal of Intelligent and Robotic Systems* 31, no. 1-3 (2001): 7-68.

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

[20] Vaghefi, A., M. A. Jafari, E. Bisse, Y. Lu, and J. Brouwer. "Modeling and forecasting of cooling and electricity load demand." *Applied Energy* 136 (2014): 186-196.

[21] Kavousi-Fard, A., and F. Kavousi-Fard. "A new hybrid correction method for short-term load forecasting based on ARIMA, SVR and CSA." *Journal of Experimental & Theoretical Artificial Intelligence* 25, no. 4 (2013): 559-574.

[22] Lin, Y., H. Luo, D. Wang, H. Guo, and K. Zhu. "An ensemble model based on machine learning methods and data preprocessing for short-term electric load forecasting." *Energies* 10, no. 8 (2017): 1186.

[23] Zheng, H., J. Yuan, and L. Chen. "Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation." *Energies* 10, no. 8 (2017): 1168.

[24] Buitrago, J., and S. Asfour. "Short-term forecasting of electric loads using nonlinear autoregressive artificial neural networks with exogenous vector inputs." *Energies* 10, no. 1 (2017): 40.

[25] Vapnik, V.. *The nature of statistical learning theory*. Springer science & business media, 2013.

[26] Pellegrini, M.. "Short-term load demand forecasting in Smart Grids using support vector regression." In *2015 IEEE 1st International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)*, pp. 264-268. IEEE, 2015.

[27] Chen, Y., P. Xu, Y. Chu, W. Li, Y. Wu, L. Ni, Y. Bao, and K. Wang. "Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings." *Applied Energy* 195 (2017): 659-670.

[28] Jang, J.. "ANFIS: adaptive-network-based fuzzy inference system." *IEEE transactions on systems, man, and cybernetics*23, no. 3 (1993): 665-685.

[29] Yang, Y., Y. Chen, Y. Wang, C. Li, and L. Li. "Modelling a combined method based on ANFIS and neural network improved by DE algorithm: A case study for short-term electricity demand forecasting." *Applied Soft Computing*49 (2016): 663-675.

[30] Çevik, H. H., and M. Çunkaş. "Short-term load forecasting using fuzzy logic and ANFIS." *Neural Computing and Applications* 26, no. 6 (2015): 1355-1367.

[31] Papaioannou, G., C. Dikaiakos, A. Dramountanis, and P. Papaioannou. "Analysis and modeling for short-to medium-term load forecasting using a hybrid manifold learning principal component model and comparison with classical statistical models (SARIMAX, Exponential Smoothing) and artificial intelligence models (ANN, SVM): The case of Greek electricity market." *Energies*9, no. 8 (2016): 635.

[32] Singh, A. K., S. Khatoon, M. Muazzam, and D. K. Chaturvedi. "Load forecasting techniques and methodologies: A review." In *2012 2nd International Conference on Power, Control and Embedded Systems*, pp. 1-10. IEEE, 2012.

[33] Massidda, L., and M. Marrocu. "Decoupling Weather Influence

from User Habits for an Optimal Electric Load Forecast System." *Energies* 10, no. 12 (2017): 2171.
[34] Kheirkhah, A., A. Azadeh, M. Saberi, A. Azaron, and H. Shakouri. "Improved estimation of electricity demand function by using of artificial neural network, principal component analysis and data envelopment analysis." *Computers & Industrial Engineering* 64, no. 1 (2013): 425-441.
[35] Yeom, C., and K. Kwak. "Short-term electricity-load forecasting using a TSK-based extreme learning machine with knowledge representation." *Energies*10, no. 10 (2017): 1613.
[36] Karunasinghe, D. S., and S. Liong. "Chaotic time series prediction with a global model: Artificial neural network." *Journal of Hydrology* 323, no. 1-4 (2006): 92-105.
[37] Oliveira, A. L., and S. R. Meira. "Detecting novelties in time series through neural networks forecasting with robust confidence intervals." *Neurocomputing* 70, no. 1-3 (2006): 79-92.
[38] Massana, J., C. Pous, L. Burgas, Joaquim Melendez, and Joan Colomer. "Short-term load forecasting in a non-residential building contrasting models and attributes." *Energy and Buildings* 92 (2015): 322-330.
[39] Dong, Y., J. Wang, C. Wang, and Z. Guo. "Research and Application of Hybrid Forecasting Model Based on an Optimal Feature Selection System—A Case Study on Electrical Load Forecasting." *Energies* 10, no. 4 (2017): 490.
[40] Son, H., and C. Kim. "Forecasting short-term electricity demand in residential sector based on support vector regression and fuzzy-rough feature selection with particle swarm optimization." *Procedia engineering* 118 (2015): 1162-1168.
[41] Burger, E. M., and S. J. Moura. "Gated ensemble learning method for demand-side electricity load forecasting." *Energy and Buildings* 109 (2015): 23-34.
[42] Jovanović, R. Ž., A. A. Sretenović, and B. D. Živković. "Ensemble of various neural networks for prediction of heating energy consumption." *Energy and Buildings* 94 (2015): 189-199.
[43] De Felice, M., A. Alessandri, and F. Catalano. "Seasonal climate forecasts for medium-term electricity demand forecasting." *Applied Energy* 137 (2015): 435-444.
[44] Hassan, S., A. Khosravi, and J. Jaafar. "Examining performance of aggregation algorithms for neural network-based electricity demand forecasting." *International Journal of Electrical Power & Energy Systems* 64 (2015): 1098-1105.
[45] Box, G. E.. "GM Jenkins Time Series Analysis: Forecasting and Control." *San Francisco, Holdan-Day* (1970).
[46] Sakia, R. M. "The Box-Cox transformation technique: a review." *Journal of the Royal Statistical Society: Series D (The Statistician)* 41, no. 2 (1992): 169-178.
[47] Azadeh, A., and V. Ebrahimipour. "An integrated approach for assessment and ranking of manufacturing systems based on machine performance." *International Journal of Industrial Engineering: Theory, Applications and Practice* 11, no. 4 (2004): 349-363.
[48] Matijaš, M., J. A. Suykens, and S. Krajcar. "Load forecasting using a multivariate meta-learning system." *Expert systems with applications* 40, no. 11 (2013): 4427-4437.