A Two-Layer Approach Model for Industry Electricity Demand in Malaysia

1. Professor Dr. Noor Azina Ismail
nazina@um.edu.my
Department of applied statistic
Faculty of Economics & Administration, University of Malaya
Kuala Lumpur, Malaysia

2. Syamnd Mirza Abdullah
cyamand@siswa.um.edu.my
Department of applied statistic
Faculty of Economics & Administration, University of Malaya
Kuala Lumpur, Malaysia

Abstract: *The main goal of this study is to propose a novel two layer approach (hybrid approach) as electricity prediction model for Malaysia. The work focuses more on the consumption rate of electricity in the industrial sector. The new hybrid approach combines a linear approach (Multiple linear regressions) and a nonlinear approach (Back Propagation Artificial Neural Network) with principal component analysis. The time span of the input data set that fed to this new proposed model is from 1992 to 2013. The included independent variables of this study are population size (Pop), Gross Domestic Product (GDP), Gross National Product (GNP), export, Import and price of electricity. The proposed combination approach can address the accuracy and reliability of the electricity demand prediction models. The Mean Absolute Percentage Error (MAPE) is used as performance indicator to evaluate the accuracy of the new hybrid approach. After testing the model, the obtained accuracy rate is 0.97. The new model used to predict the industrial demand rates for next five years (2015 – 2020).*

Keywords: *electricity prediction, Industry sector in Malaysia, Two-Layer Approach, principal component regression, back Propagation Neural Network.*

1. Introduction

The energy of electric plays an important role for technical, social and economic development in any country. Most human live activities, such as industrial, agriculture, transportation, lighting and heating depend on the energy of electricity [1-3]. Increasing electricity consumption per capita is characterized as the lifestyle of citizens in that country. Therefore, it is very important to do planning and supporting the policy maker on consumption and distribution strategies of electricity. Such steps could be done through analyzing and identifying the development of energy prediction [4]. Typical consumption of electricity has an increasing rate. Therefore, the electricity demand prediction models is very important, as the product rate of supplying electric plants is based on that prediction [1].

Malaysia's net electricity demand has increased at an average rate of 59.37% annually from 2000 to 2014. The consumption of electricity in 1992 was 2218 ktoes, and this amount has increased to 10011 ktoes in 2012. As illustrates in Fig 1, most growth in consumption of electricity in Malaysia has been recorded in the industrial sector. This is because the number of applications intensive electricity of technologies has been increasing in the industrial sector. The industrial sector of electricity consumption has increased from 1137 ktoe in 1992 to 4509 ktoe in 2014. It means the annual average rate of the growth of electricity consumption in the industrial sector was 46.6% between 2000 and 2013 [5].

The electricity demand prediction can be classified into three types; short term, medium term and long term and they range from one hour to one week, one month to one year and one year for decades respectively[6]. This study used a long-term prediction model, because so far, less study is available in the literature considering long term in Malaysian industrial sector to predict model.

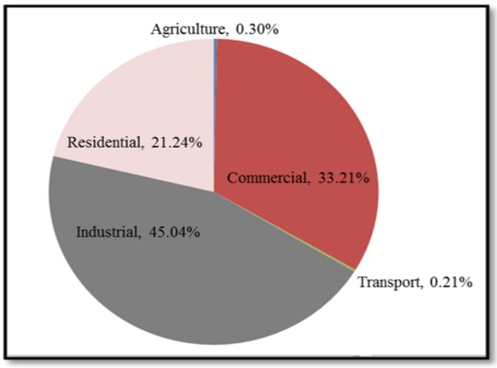

Fig 1. Electricity consumption behaviour for different Sector in Malaysia (2012 year)

Currently, most techniques are frequently employed to predict electricity demand were regression model and artificial intelligent. Multivariate linear regression (MLR) was utilized to develop prediction electricity demand of the Jordanian industrial sector[7]. Prediction of Turkey's industrial sector electricity demand, Genetic Algorithm[8], Artificial neural network (ANN)[9] and structural time series of analysis[10]. Many studies have been conducted on the prediction of the electricity demand model [2, 11-14]. The Principal component regression with back-propagation artificial neural network (PCR-BPANN) approach is still not explored in the industrial prediction of electricity demand.

2. Influential Factors

There are several important factors are affected on the prediction electricity consumption of industrial in long term such as;

- Economic factors: The increase of the economic growth is directly related to the increase the energy consumption, especially in the electricity consumption, the Gross Domestic Product (GDP) and Gross National Product (GNP) are measures for activities of economic [9]. The R^2 value between GDP and industrial electricity consumption is 0.94 and GNP with industrial electricity consumption is 0.95.
- Import and Export: The import and export related to increase or decrease electricity consumption of industrial sector. The R^2 value between import with electricity consumption of industrial is 0.92 and export with electricity consumption of industrial is 0.94
- Electricity price: it has effect on increase or decrease the electricity demand such as in Malaysia[16]. TheR^2 value = 0.61
- Population: The statistical correlation coefficient of multiple determinations (R^2 value) for industrial electricity consumption vs. Population is 0.71.

3. Methodology

3.1. Data Collected

To collect required data on industrial electricity consumption rates and records of significant independent variables, this study consulted the department of "Statistic in Malaysia" to get records on Malaysia (Malaysia Economic statistic time series, 2011). The data recorded on the industrial electricity consumption (Kote) and influenced by many factors: population size, Gross Domestic Product (GDP), Gross National Product (GNP), export, Import and price of electricity. The annually historical data related these factors were collected from 1992 to 2013.

3.2. The Two-Layer Model

The selected factors that mentioned in section-2 have been fed to the proposed model. The model structured using multiple linear regression (MLR) with back-propagation neural network (BPNN) based on the principal component analysis (PCA). The model predicts long-term electricity consumption for industrial sector in Malaysia. The first part of this long term based model is combining PCA with MLR to form Principle Component Regression (PCR).

PCA was first proposed by Hotelling in 1933. PCA is a multivariate statistical technique widely employed in data analysis in different areas, because it is very simple, nonparametric method. The main objective of PCA is to reduce the number of independent variables and transform them into new uncorrelated variables that called principal component (PCs).Eigen analysis is a mathematical process in PCA. Equation -1 is used to compute the Eigen value in a standardized matrix format. Basically, the Eigen vectors related with the largest Eigen values have the same direction as the first PC. The Eigen vectors related with the second largest Eigen values determines the direction as the second PC.

$$|C - \gimel\, I| = 0 \qquad (1)$$

Where *C* is the correlation matrix of standardized dataset, $\gimel$ is the eigenvalue and *I* is the identify matrix.

The weight of the variables in the PC is then obtained by equation (2).

$$|C - \gimel\, I|W = 0 \qquad (2)$$

Where *W* is the matrix of the weights.

To assess the influences of each of the independent variable in the PC, varimax rotation was used to get values of rotated factor loadings. These loading show the contribution of each independent variable in a certain PC.

The PC employed for the prediction of industrial electricity consumption was obtained during multiplication of the standardized dataset matrix by the previously computed weight (*W*).

To apply the PCA technique of the dataset, this study was verified during the tests of modified KMO and Bartlett's of sphericity using equation (3).

$$X_k^2 = \left[n - k - \frac{2(p-k) + 7 + \frac{2}{(p-k)}}{6} + \sum_{j=1}^{k} \left(\frac{\bar{\lambda}}{\lambda_j - \bar{\lambda}}\right)^2 \right] * \left[-\ln \prod_{j=k+1}^{p} \lambda_j + (p-k) \ln \bar{\lambda} \right] \quad (3)$$

Where n the number of the observations dataset, p is the number of components, λ_j illustrates the eigenvalue for the $k_j th$ components.

However the value of $\bar{\lambda}$ is computed by equation (4):

$$\bar{\lambda} \sum_{j=k+1}^{p} \frac{\lambda_j}{p-k} \quad (4)$$

The null hypothesis is to determine that all independent variables were uncorrelated among them and when accepted, PCA could be applied [17].

Therefore, this study conducted the PCA statistical technique to remove or reduce multicollinearity problem among independent variables. One advantage of PCA technique is transforming the original dataset to combine linear uncorrelated dataset. This combined linear datasets can be fed to any prediction models as a new dataset. However, the new dataset still explains most information and records in the original dataset. The optimal number of PCs can be determined through each eigenvalue. Principal components can be ranked according to their ability to explain variance in the original data set.

The second part of the PCR is MLR technique. This technique is employed for modeling and analysis of independent variables, and can be used to determine which independent variable has the most impact on the variance of the dependent variable. However, MLR model are also used as prediction model. The general Equation of MLR as prediction is shown in equation-5.

$$\dot{Y} = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n \quad (5)$$

Where: Y' is an estimating for the output electricity consumption industrial sector. However, β_1 and β_2 are standardized coefficients of the independent variables and $X_1 \ldots\ldots X_n$ are the independent variables (predictors).

With PCR, the problems of multicollinearity among independent variables and dimensionality size of input dataset have been solved. However, due to linearity respond of PCR in fitting with predictors, a problem of residual errors is raised, and this problem affects the accuracy of the proposed prediction model.

To improve such accuracy problem, this study proposes ANN, which is nonlinear approach. This approach is used to overcome the residual errors that recorded from the linear part (PCR) results.

ANN provides best results for nonlinear dataset among input variables [21]. One of the most widely applied algorithms in neural network models is the back propagation algorithm (BP), which is more accurate than other algorithms in obtaining minimum errors in the ANN applications[18]

The architecture of neural network consists of three layers such as input, hidden and output layers, as illustrated in fig (2). Circles in figure (2) represent the neurons and are arranged in mentioned three layers. The input layer contains only one layer as output layer, but the hidden layer may be more than one layer with respect to the complexity of the received information[19]. As mentioned before, this NN is employed to solve the residual error problem. Therefore, the output function of this neural network will be formulated as in equation-6.

$$e_t = f(e_{t-1}, e_{t-2}, \ldots\ldots\ldots e_{t-n}) + \epsilon_t \quad (6)$$

where e_t is the residual error at the time *t* from the PCR model and f is nonlinear function that determined by neural network. While, ϵ_t is the randomly of error.

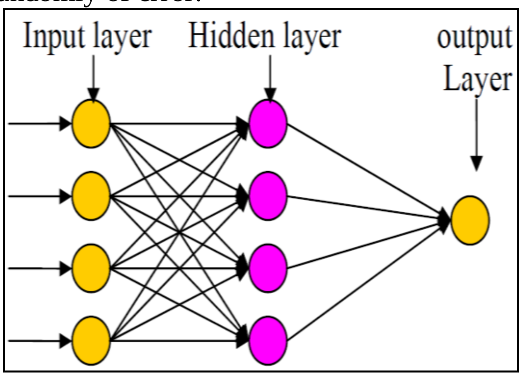

Fig 2. The architecture of ANN.

The results of prediction model from the BPNN model is denoted by $\tilde{\varepsilon}_t$, and the combined prediction will be shown by equation (7).

$$\tilde{K} = \hat{Y} + \tilde{\varepsilon}_t \qquad (7)$$

Where $\tilde{K}$ is the hybrid approach that estimated by PCR-BPNN model, $\hat{Y}$ is linear component that estimated by PCR and $\tilde{Z}_t$ is the nonlinear approach BPNN that estimated residual error from PCR model.

The main idea of this combining is to capture different patterns of the data set by using the different features model. Therefore, the accuracy of such hybrid prediction model is always better than the individual prediction model[20]. This combination in electricity prediction model (principal component regression with back-propagation artificial neural network) is new, based on the best knowledge of the authors. The proposed model can capture both linear and nonlinear dataset, and can solve multicollinearity and dimensionality problems of input dataset of. Such kind model still unexplored in the demand prediction of the industrial sector in Malaysia. Figure 3 shows the structure of this hybrid combination.

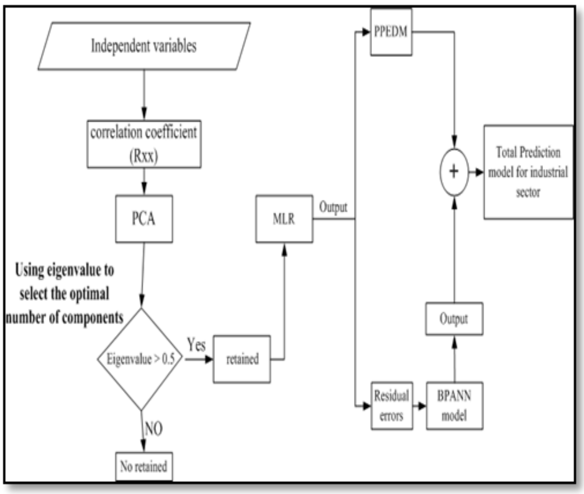

Fig 3: prediction of Malaysia's industrial sector

3.3. Performance index

The last step of this work is evaluating the prediction rates and errors of the proposed model. There are several ways of measuring the performance indicators of the prediction model. However, in this study only two important measurements have been focused; mean absolute percentage error (MAPE) and Root mean square error (RMSE). The formula of each of them has shown in the equations 8 and equation 9.

$$MAPE = \left(\frac{1}{n}\sum_{i=0}^{n}\left|\frac{Y_i - Y_i'}{Y_i}\right|\right) \times 100 \qquad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=0}^{n}(Y_i - Y_i')^2}{n}} \qquad (9)$$

Where Y_i, Y_i' were the actual values of electricity consumption for industrial sector in year i and estimated industrial electricity consumption which computed by formula (7), and 'n' denoted of the number years, which is 23.

4. Result and Discussion

This study applied standardization method on the original dataset using equation (10). The process of standardization is essential to put records in the dataset in one measure or scale.

$$x' = \frac{x - \bar{x}}{\sigma} \qquad (10)$$

Where x' is the standardization dataset, for each original independent variables and dependent variable, $\bar{x}$ is the average of the original dataset and σ is the standard deviation of original dataset.

The standardized dataset is processed to find correlation among independent variable. Table 2, shows the results of this correlation coefficients. The result showed that the correlation among independent variables (population, GDP, GNP, export and import and electricity price is more than 0.5. Such results mean that high correlations among them are existing, which in turn means occurrence of multicollinearity problem. Therefore, all variables have been fed to PCA technique in order to reduce or remove multicollinearity problem and to get a new uncorrelated dataset.
This work uses two types of statistical test (KMO test and Bartlett's Test) to measure the sampling adequacy. The accepted value of KMO for any dataset should be more than 0.6, while the accepted value of the Bartlett's test is less than 0.05. As shown in table 3, the result of the KMO test is 0.737 which is more than 0.5. While, for the Bartlett's Test of sphericity the result is 0.000, which is less than 0.05.

Therefore, results of both tests that shown in table-3 are supporting the necessity of applying PCA on the present dataset.

Table 2: correlation coefficients among independent variables

	Pop	GNP	GDP	Export	import	electricity Price
Pop	1.00	0.635	0.658	0.610	0.575	0.608
GNP	0.635	1.00	0.997	0.992	0.982	0.494
GDP	0.658	0.997	1.00	0.993	0.984	0.535
Export	0.610	.992	.993	1.00	0.992	0.506
import	.575	0.982	0.984	0.992	1.00	0.499
electricity Price	0.608	0.494	0.535	0.506	0.499	1.00

Table 3: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.720
Bartlett's Test of sphericity	Approx. Chi-Square	291.039
	df	15
	Sig.	.000

The optimal number of PCs is considered only when the eigenvalues equal to or more than one with at least 80% cumulative variance. In this study only the first PC is greater than one, however the cumulative variance is 75.444. The second PC that so close to one is 0.868. Therefore, this study selected PC_1 and PC_2 to increase the explained cumulative variance. Table 4 shows a summary of the six independent variables for PCA and Figure 1 presents the result of the number principal component with eigenvalues.

The equation of multiple linear regression model that combined with PC_1 and PC_2 as independent variables is shown in equation-11.

$$Y' = \beta_1 PC_1 + \beta_2 PC_2 \qquad (11)$$

Where: Y' is an estimating for the output electricity consumption industrial sector. However, β_1 and β_2 are standardized coefficients of the first and second PCs respectively.

Table 4: initial and extracted independent variables

	Initial Eigenvalue		
Component	**Total**	**Variance (%)**	**Cumulative (%)**
PC1	4.527	75.444	75.444
PC2	.868	14.463	89.908
PC3	.581	9.688	99.596
PC4	.018	.304	99.900
PC5	.005	.079	99.978
PC6	.001	.022	100.000

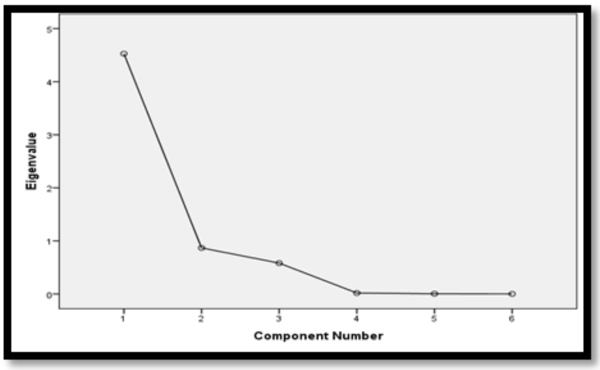

Fig. 1 the result of the eigenvalue and optimal number of PCs

Fig.2 presents the comparison between the curves of the actual electricity consumption and consumption rates that predicted by the multiple linear regression model.

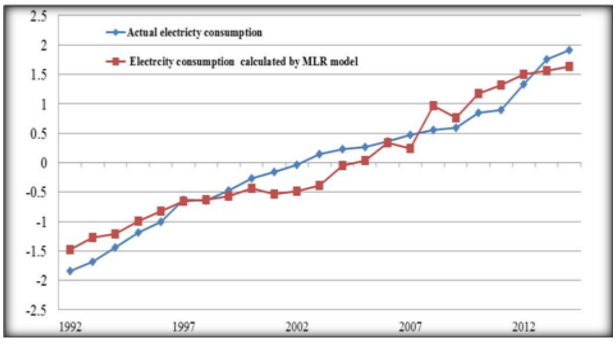

Fig.2 the actual and calculated electricity consumption

As the linear model (MLR) can't capture nonlinear pattern of the dataset, therefore the residual errors of the MLR model are recorded as shown in the figure-2. The result from back-Propagation Neural network model can be employed as a prediction model for the error rates of the MLR model. Fig 3 shows the result of the hybrid approach with the actual dataset. The result shows that actual data and predicted resulted becomes more closer than shown in finger-2. The new hybrid system fits the actual and predicted results more than other techniques can do. To check this, another test has been evaluated.

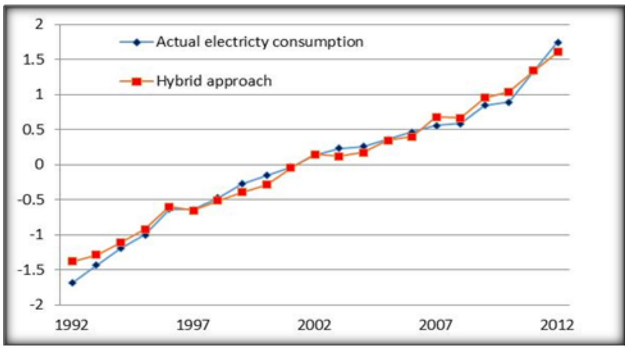

Fig 3: Hybrid approach (PCR-BPANN)

To achieve this test, this work prefixes PCA to BPNN to form PC-BPNN hybrid system. The model can analyse the nonlinear pattern of any dataset, however, it faces some difficulties with linear patterns. As shown in the figure-4, the actual dataset and predicted are fitted, but not like the fit shown in figure-3.

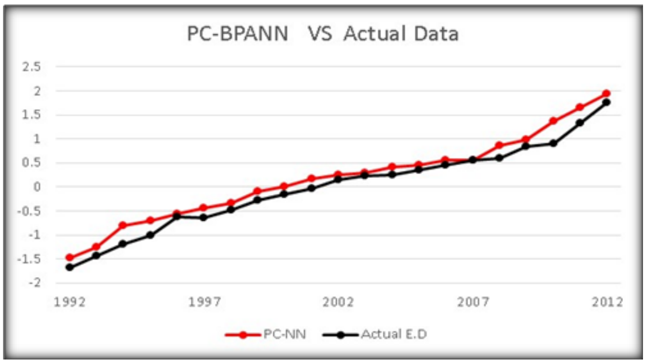

Fig 4: PC – BPANN model

Table 5 presents some indicators to evaluate the performance of the PCR, PC-BPANN and PCR-BPANN models. This evaluation covers the R^2, MAPE, and RMSE indicators. Results show that PCR-BPNN is outperformed than other types of hybrid models of prediction.

Table 5: result of MLR model

Models	**R^2**	**MAPE %**	**RMSE**
PCR	0.92	21.9	0.29
PC-BPANN	0.95	18.1	0.18
PCR-BPANN	0.97	8.821	0.11

5. Conclusion

- The study conducted PCA statistical technique of sex prevailing factors the electricity demand. The first and second PCs had eigenvalue close to and more than one and cumulative percentage variance of 75.4% and 89.91%, respectively.
- The hybrid approach (PCR-BPANN) was utilized to calculate electricity consumption for industrial sector in Malaysia. The independent variables that selected from first and second principal components of the PCA. The R-square was determined is (0.97) for predicted and actual data of electricity consumption. The obtained error rate showed that the PCR-BPNN was so close to calculate estimation electricity consumption with the actual electricity consumption.
- The values of RMSE and MAPE show that the proposed model PCR-BPNN is more stable, accruable and reliable than other major prediction models.

Acknowledgment:

I would like to than University of Malaya, Faculty of Economic and Administration, Department of Applied Statistic for providing all supports to achieve this project as a part of PhD requirement. Moreover, Technical Institute of Koya deserved thanks for their support in this PhD project.

References

[1] Bilgili, M., et al., Electric energy demands of Turkey in residential and industrial sectors. Renewable and Sustainable Energy Reviews, 2012. 16(1): p. 404-414.

[2] Saravanan, S., S. Kannan, and C. Thangaraj, Prediction of India's Industrial Sector Electricity Consumption Using ANFIS, in Power Electronics and Renewable Energy Systems. 2015, Springer. p. 1281-1288.

[3] Akdemir, B. and N. Çetinkaya, Long-term load forecasting based on adaptive neural fuzzy inference system using real energy data. Energy Procedia, 2012. 14(0): p. 794-799.

[4] Azadeh, A. and Z.S. Faiz, A meta-heuristic framework for forecasting household electricity consumption. Applied Soft Computing, 2011. 11(1): p. 614-620.

[5] Saravanan, S., et al., Forecasting India's Electricity Consumption Using Particle Swarm Optimization, in Power Electronics and Renewable Energy Systems, C. Kamalakannan, et al., Editors. 2015, Springer India. p. 843-851.

[6] Çunkaş, M. and A. Altun, Long term electricity demand forecasting in Turkey using artificial neural networks. Energy Sources, Part B: Economics, Planning, and Policy, 2010. 5(3): p. 279-289.

[7] Al-Ghandoor, A., et al., Electricity consumption and associated GHG emissions of the Jordanian industrial sector: Empirical analysis and future projection. Energy Policy, 2008. 36(1): p. 258-267.

[8] Ozturk, H.K. and H. Ceylan, Forecasting total and industrial sector electricity demand based on

genetic algorithm approach: Turkey case study. International journal of energy research, 2005. 29(9): p. 829-840.
[9] Zahedi, G., et al., Electricity demand estimation using an adaptive neuro-fuzzy network: a case study from the Ontario province–Canada. Energy, 2013. 49: p. 323-328.
[10] Dilaver, Z. and L.C. Hunt, Industrial electricity demand for Turkey: a structural time series analysis. Energy Economics, 2011. 33(3): p. 426-436.
[11] Kandananond, K., Forecasting electricity demand in Thailand with an artificial neural network approach. Energies, 2011. 4(8): p. 1246-1257.
[12] Kavaklioglu, K., Modeling and prediction of Turkey's electricity consumption using Support Vector Regression. Applied Energy, 2011. 88(1): p. 368-375.
[13] Kavaklioglu, K., et al., Modeling and prediction of Turkey's electricity consumption using Artificial Neural Networks. Energy Conversion and Management, 2009. 50(11): p. 2719-2727.
[14] Suganthi, L. and A.A. Samuel, Energy models for demand forecasting—A review. Renewable and Sustainable Energy Reviews, 2012. 16(2): p. 1223-1240.
[15] Kargar, M.J. and D.K. Charsoghi, Predicting Annual Electricity Consumption In Iran Using Artificial Neural Networks (Narx). Indian J. Sci. Res, 2014. 5(1): P. 231-242.
[16] Pourazarm, E., Electricity demand analysis in different sectors: a case study of Iran. 2012.
[17] Peres-Neto, P.R., D.A. Jackson, and K.M. Somers, How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Computational Statistics & Data Analysis, 2005. 49(4): p. 974-997.
[18] Li, J., et al., Brief Introduction of Back Propagation (BP) Neural Network Algorithm and Its Improvement, in Advances in Computer Science and Information Engineering. 2012, Springer. p. 553-558.
[19] Daliakopoulos, I.N., P. Coulibaly, and I.K. Tsanis, Groundwater level forecasting using artificial neural networks. Journal of Hydrology, 2005. 309(1): p. 229-240.
[20] Zhang, G.P., Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 2003. 50: p. 159-175.