

# Input Variable Selection for RBFN-based Electric Utility's CO<sub>2</sub> Emissions Forecasting

I. Falconett and K. Nagasaka

**Abstract**—This study investigates the performance of radial basis function networks (RBFN) in forecasting the monthly CO<sub>2</sub> emissions of an electric power utility. We also propose a method for input variable selection. This method is based on identifying the general relationships between groups of input candidates and the output. The effect that each input has on the forecasting error is examined by removing all inputs except the variable to be investigated from its group, calculating the networks parameter and performing the forecast. Finally, the new forecasting error is compared with the reference model. Eight input variables were identified as the most relevant, which is significantly less than our reference model with 30 input variables. The simulation results demonstrate that the model with the 8 inputs selected using the method introduced in this study performs as accurate as the reference model, while also being the most parsimonious.

**Keywords**—Correlation analysis, CO<sub>2</sub> emissions forecasting, Electric power utility, Radial basis function networks.

## I. INTRODUCTION

THE need to reduce green house gases emissions will impose a heavy burden on the world economy. Since the electric power industry is an important direct and indirect source of CO<sub>2</sub> emissions, the countermeasures against global warming are expected to have a great impact on the electricity sector. For instance, in the fiscal year (FY) 2008 the estimated CO<sub>2</sub> emissions in Japan resulting from the electricity use were about 395 million tons of CO<sub>2</sub> which represents approximately 30 % of the total CO<sub>2</sub> Japanese emissions.

Furthermore, nowadays the regulatory authorities need ahead information about the CO<sub>2</sub> emissions from power generating companies, in order to elaborate an accurate allocation plan or establish a carbon taxes with small welfare losses. On the other hand, the electric power utilities should be aware of the CO<sub>2</sub> emissions reduction costs for the next year as well as how many emissions permit they need to purchase through the emission trading scheme to accomplish their imposed targets. Therefore, producing optimal CO<sub>2</sub> emissions forecasts are important in order to implement the necessary strategies to tackle global warming.

In recent years, many approaches have been developed for

calculating the CO<sub>2</sub> emissions forecast on a global and country level. In particular, Kaimuna et al. (2000) [1] applied AIM model for forecasting the CO<sub>2</sub> emissions in Japan. The AIM model is a computable general equilibrium model which estimates the greenhouse gas emissions and assesses policy options to reduce them in the Asian-Pacific region. Another computable general equilibrium model that has been widely used to analyze the emission of greenhouse gases is MIT EPPA model. Kasahara et al. (2007) [2] employed MIT EPPA model to project the economic growth and simulated the CO<sub>2</sub> emissions of Japan up to the year 2020. Yang and Schneider (1998) [3] applied a statistical decomposition model to analyze different carbon dioxide emissions scenarios in three regions: more developed countries, China and the remaining less developed countries. In Yang and Schneider approach, CO<sub>2</sub> emissions are decomposed into the product of four factors: population size, affluence (GDP per capita), energy intensity (energy use per unit GDP) and carbon intensity (CO<sub>2</sub> emissions per unit energy). Schmalensee et al. (1998) [4] and Holtz-Eakin and Selden (1995) [5] used the Environmental Kuznet curve to construct a CO<sub>2</sub> emissions business as usual scenarios. These approaches are based on the relationship between emissions per capita and income. In spite of the fact that the CO<sub>2</sub> emissions forecasts on a country level are calculated using a variety of models, there is scarce literature on the CO<sub>2</sub> emissions prediction from the electric power companies. For instance, Islam et al. (1997) [6] applied a linear time series analysis to predict the CO<sub>2</sub> emissions of an electric utility. In order to contribute with the CO<sub>2</sub> emissions forecasting literature, we propose an artificial neural networks (ANNs) model, specifically radial basis function networks (RBFN), to forecast the CO<sub>2</sub> emissions of an electric power utility. Based on the fact that neural networks are able to provide good solutions to the problem of modeling complex nonlinear relationships [7] – [9].

In addition, the second goal of this paper is the important issue of input variable selection for RBFN-based electric utility's CO<sub>2</sub> emissions forecasting. The importance of this problem is evident because of ANNs learn the relationships between input and output variables on the basis of provided input-output data pairs. Therefore, if correlated or insignificant variables are selected as inputs, bigger training sets are required, then computational resources are wasted during training and sometimes inferior results are obtained.

I. Falconett is with Department of Electronic and Information Engineering, Tokyo University of Agriculture and Technology, 2-24-16 Naka-cho, Koganei-shi, Tokyo 184-8588, Japan (e-mail: irigi@hotmail.com).

K. Nagasaka is with Department of Electronic and Information Engineering, Tokyo University of Agriculture and Technology, 2-24-16 Naka-cho, Koganei-shi, Tokyo 184-8588, Japan (e-mail: bahman@cc.tuat.ac.jp).

## II. MODEL FORMULATION

ANN methods are good choice to predict the CO<sub>2</sub> emissions of an electric utility, since these techniques do not require a priori postulation of model parameters to represent the complex relationship between the CO<sub>2</sub> emissions and the factors that determine it. In fact, the model parameters are iteratively adjusted and optimized through network learning of historical patterns. And, the time series forecasting is performed completely by inference of future behavior from example of past behavior.

### A. Network Architecture

A RBFN was used here to forecast the monthly CO<sub>2</sub> emissions of Tokyo Electric Power Company. The architecture of RBFN consists of three layers as follows,

#### 1) The Input Layer

The input layer is made up of source nodes whose number is equal to the dimension of the input vector  $x_t$ . For  $n$  measurements from an output  $y_t$  and  $d$  inputs, the input vector is thus represented by  $x_t = [x_{t1}, \dots, x_{td}]$ ,  $t = 1$  to  $n$ . Besides, the data at the input layer were normalized to mean zero and standard deviation 1 to make the relevance measures comparable. Then, the input neurons feed the values to each of the neurons in the hidden layer.

#### 2) The Hidden Layer

The hidden layer consists of a set of basis function units that carry out a nonlinear transformation from the input space to the hidden space. Usually, the nonlinear transformation is based on Gaussian function which characteristic feature is that its response decreases monotonically with the distance from a central point. The activation function for a radial basis neuron is,

$$\Phi(x_t, c_i, \sigma_i) = e^{-\left(\frac{\|x_t - c_i\|^2}{2\sigma_i^2}\right)} \quad t = 1 \text{ to } n \quad (1)$$

Where  $\|\dots\|$  represents the Euclidean distance,  $c_i$ ,  $\sigma_i$  and  $\Phi$  are the centers, the width of the basis function and the output of the hidden unit  $i$ , respectively [10].

The center of the basis function for a node  $i$  at the hidden layer is a vector  $c_i$  whose size is the same as the input vector  $x_t$ . The position of the center is chosen according to the distribution of input variables in space. At locations where there are few inputs few nodes will be placed and conversely, a lot of nodes will be placed where there are many input data. Further, the optimal number of center is determined by the training process.

In this study, a trial and error procedure for the width parameter ( $\sigma_i$ ) selection was used by gradually varying the value of the width parameter. Moreover, we choose to work with the same  $\sigma_i$  for all hidden units. Next, the resulting value for the radial basis function is passed to the summation layer.

#### 3) The Output Layer

The value from each neuron in the hidden layer is multiplied by a weight associated with the neuron ( $w_1, w_2, \dots, w_p$ ) and passed to the summation which adds up the weighted values and presents this sum as the output of the network:

$$\hat{y}_t = \sum_{i=1}^p w_i \Phi(x_t, c_i, \sigma_i) \quad (2)$$

The computation of the optimal weight values between the neurons in the hidden layer and the summation layer is done by using least squares method. Since we applied a supervised learning, our training data set consist of pairs of inputs and target outputs. Thus, knowing both the network outputs and the target outputs, the error is minimized by automatically adjusting the weights.

### B. Performance Criteria

The evaluation of forecasting capability of developed RBFN model is examined by using root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). The RMSE is the square root of the average of square differences between the forecasted CO<sub>2</sub> emissions  $\hat{y}_t$  and the actual ones  $y_t$ :

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (3)$$

The MAE describes the average magnitude of the errors without considering their direction. It is computed and given as:

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (4)$$

These two errors indicators show the errors in the same units and scale as the parameter itself. The MAE and RMSE can be used together to diagnose the variation in the errors. A large difference between the MAE and RMSE reveals a large variation in the error time series. Besides, the RMSE is larger or equal than MAE.

In order to assess the prediction accuracy of RBFN model as a percentage, we applied MAPE which is calculated using the following definition:

$$MAPE = \frac{100}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{y_t} \quad (5)$$

## III. INPUT VARIABLE SELECTION

In a RBFN framework a key issue is the choice of a proper topology for a given problem. This requires determining the RBFN parameter as well as careful selection of the appropriate input variables. The RBFN parameters that affect the performance of the model are: the number of hidden units or centers, the position of the centers, the width of the radial basis function and the weights from the hidden layer to the final network layer. These network parameters were found out during the training process. As for the input selection, the applied procedure is depicted in Fig. 1 and explained below,

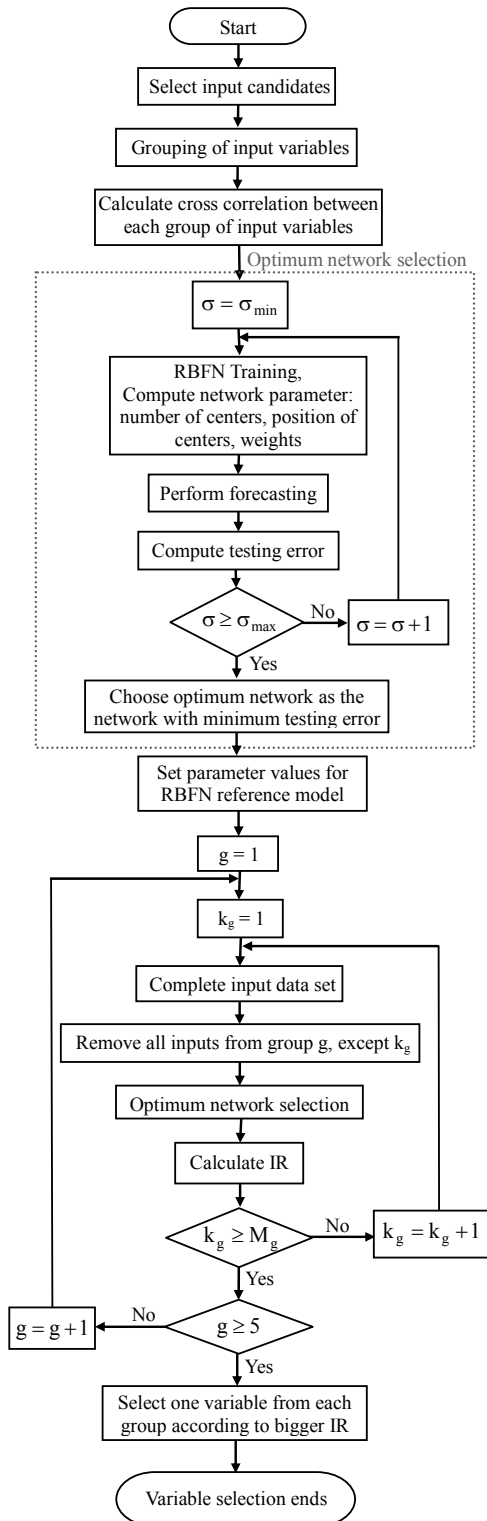


Fig. 1 Flowchart of the proposed method for input variable selection.  
 $k_g$  = Variable to be investigated from group  $g$   
 $M_g$  = Maximum number of variables in the group  $g$

1) Selection of input variables for the reference model

A set of input variables that are significantly correlated with the output variable were selected from available data. In this study, 29 input variables and 1 seasonal parameter (month of the year) were employed to implement our RBFN reference model as it is shown in Fig. 2.

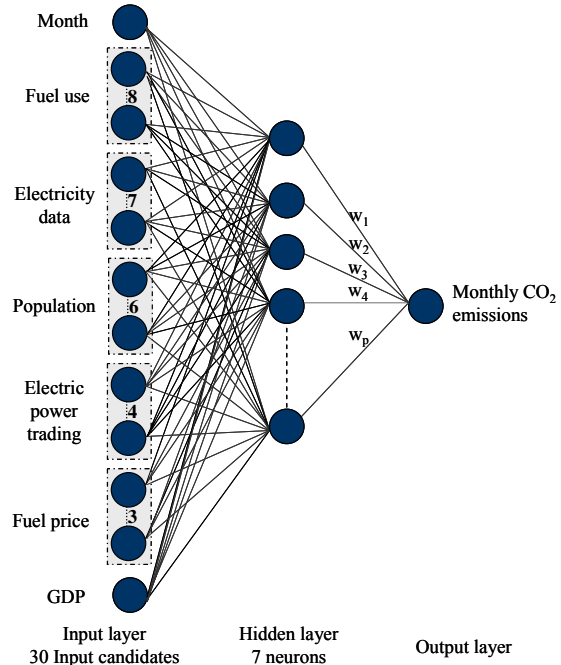


Fig. 2 Architecture of RBFN reference model.

2) Grouping of input variables

These 29 input variables were divided in 6 groups which describe the principal characteristics of these input candidates (see Table I). These groups of variables are considered in the literature to be the major determinants of CO<sub>2</sub> emissions level.

The historical data (FY 2005 - FY 2008) came from several sources. Specifically, the fuel use data and the electricity data were obtained from Federation of Electric Power Companies of Japan (FEPC). The population data came from the Ministry of Internal Affairs and Communication Statistic bureau. The energy trading data were based on the information of Japan Electric Power Exchange (JPEX). The economic indicator GDP came from Department of National Accounts, Economic and Social Research Institute (ESRI). The source of the fuel price data was the Energy Data and Modeling Center, The Institute of Energy Economics, Japan. For estimating CO<sub>2</sub> emissions from electricity generation, data on CO<sub>2</sub> emission intensity and CO<sub>2</sub> emissions level were based on several reports from Tokyo Electric Power Company and FEPC.

3) Correlation analysis

A correlation analysis between variables inside of each group was carried out (see Tables II, III, IV, V and VI), in order to facilitate the selection of the most representative variable from each group.

TABLE I  
CORRELATION BETWEEN CO<sub>2</sub> EMISSIONS AND THE INPUT CANDIDATES

Variable	Symbol	CO <sub>2</sub> Emissions (Million ton)
<b>Group 1: Fuel use</b>		
Coal purchased (t)	F1	0.251
Coal consumed (t)	F2	0.389
Heavy oil purchased (kL)	F3	0.469
Heavy oil consumed (kL)	F4	0.456
Crude oil purchased (kL)	F5	0.373
Crude oil consumed (kL)	F6	0.285
LNG purchased (t)	F7	0.482
LNG consumed (t)	F8	0.513
<b>Group 2: Electricity data</b>		
Hydro generation (MWh)	E1	-0.048
Thermal generation (MWh)	E2	0.502
Nuclear generation (MWh)	E3	-0.224
Electricity purchased (MWh)	E4	0.594
Electricity supply (MWh)	E5	0.574
Electricity demand (MWh)	E6	0.535
Peak load (MW)	E7	0.518
<b>Group 3: Population</b>		
Whole Japan (thousand persons)	P1	-0.185
Saitama (thousand persons)	P2	0.653
Chiba (thousand persons)	P3	0.639
Ku-area (thousand persons)	P4	0.650
Yokohama (thousand persons)	P5	0.649
Kawasaki (thousand persons)	P6	0.651
<b>Group 4: Electric power trading</b>		
Trading volume (kWh)	T1	0.580
24-hour mean value price (yen/kWh)	T2	0.538
Day time price (yen/kWh)	T3	0.587
System peak price (yen/kWh)	T4	0.613
<b>Group 5: Fuel price</b>		
Coal price (US\$/t)	D1	0.675
Oil price (US\$/barrel)	D2	0.468
LNG price (US\$/m <sup>3</sup> )	D3	0.418
<b>Group 6: Economic indicator</b>		
Real GDP (billion of yen)	I1	0.304

TABLE II  
CORRELATION BETWEEN EACH INPUT CANDIDATES IN FUEL USE GROUP

Group 1	F1	F2	F3	F4	F5	F6	F7	F8
F1	1.00	0.64	0.32	0.31	0.27	0.31	0.19	0.45
F2		1.00	0.49	0.52	0.34	0.33	0.53	0.65
F3			1.00	0.96	0.84	0.78	0.70	0.83
F4				1.00	0.81	0.76	0.67	0.84
F5					1.00	0.72	0.54	0.71
F6						1.00	0.46	0.65
F7							1.00	0.77
F8								1.00

As we can see in Table III and Table IV, the electricity data group and population group were the only groups which some input variables are negative correlated. Besides, in the

population group, all candidates are highly correlated except for whole Japan population input. There is not significant relationship between hydro generation and the following variables: electricity purchased, electricity supply and electricity demand. In the same way, the correlations between nuclear generation and electricity supply, electricity demand and peak load are very low.

TABLE III  
CORRELATION BETWEEN EACH INPUT CANDIDATES IN THE ELECTRICITY GROUP

Group 2	E1	E2	E3	E4	E5	E6	E7
E1	1.00	-0.29	0.43	-0.05	0.02	0.02	0.21
E2		1.00	-0.77	0.67	0.61	0.46	0.44
E3			1.00	-0.41	-0.06	0.01	0.05
E4				1.00	0.69	0.55	0.67
E5					1.00	0.76	0.89
E6						1.00	0.71
E7							1.00

TABLE IV  
CORRELATION BETWEEN EACH INPUT CANDIDATES IN THE POPULATION GROUP

Group 3	P1	P2	P3	P4	P5	P6
P1	1.000	-0.541	-0.539	-0.536	-0.534	-0.538
P2		1.000	0.997	0.999	0.998	0.998
P3			1.000	0.998	0.998	0.999
P4				1.000	0.999	0.999
P5					1.000	0.999
P6						1.000

TABLE V  
CORRELATION BETWEEN EACH INPUT CANDIDATES IN THE ELECTRIC POWER TRADING GROUP

Group 4	T1	T2	T3	T4
T1	1.000	0.445	0.388	0.328
T2		1.000	0.960	0.882
T3			1.000	0.963
T4				1.000

TABLE VI  
CORRELATION BETWEEN EACH INPUT CANDIDATES IN THE FUEL PRICE GROUP

Group 5	D1	D2	D3
D1	1.000	0.587	0.648
D2		1.000	0.858
D3			1.000

#### 4) Training and selection of optimum network for the reference model

ANN training consists on determining the network parameters that allow achieving the desired target based on the available training data sets. The training and network selection procedure is as follows: First, the minimum width parameter ( $\sigma_{\min}$ ) and maximum width parameter ( $\sigma_{\max}$ ) are fixed. Then, the optimal width is obtained by an iterative process as it is shown in Fig. 1. For each width parameter, the RBFN trains with an increasing number of centers. And, the position of centers is determined by clustering algorithms. Next, the weights connecting the hidden layer with the output layer are determined by least mean

squared (LMS) algorithms. The training stops when the validation error measured over 5 epochs has increased more than fixed threshold since the last time it decreased or the maximum number of centers has been reached. After training stops, the model with the lowest validation set error is used to forecast the CO<sub>2</sub> emissions for the testing data set and the testing error is calculated. The iteration continues until  $\sigma_{\max}$  is reached. Finally, the optimum network is selected based on the lowest testing error and the network parameter values are set.

#### 5) Variable testing

The variable selection procedure is based on the effect that each input has on the forecasting error. The change in forecasting error is examined by removing all inputs except the variable to be investigated from its group, calculating the networks parameter and performing the forecasting. Finally, the new forecasting error is compared with the reference model.

#### 6) Improvement rate

The improvement rate (IR) is applied in order to compare the prediction error between the RBFN model under evaluation and the reference model,

$$IR = \frac{RMSE_{RM} - RMSE_N}{RMSE_{RM}} \quad (6)$$

where the subscripts RM and N of RMSE refer to the reference model and the model under study respectively.

#### 7) Selection of one representative input variable from each group

After, the improvement rates were calculated for our 29 input candidates, one variable was selected from each group based on the bigger IR. Then, these results were compared with the correlation analysis.

### IV. SIMULATION RESULTS AND DISCUSSION

This paper has two important goals: namely, the evaluation of performance of RBFN in the monthly CO<sub>2</sub> emissions forecasting of an electric power utility and identification of best input variables for RBFN CO<sub>2</sub> emissions prediction. Several interesting conclusions may be drawn from our simulation results.

#### A. Reference Model

The RBFN as shown in Fig. 2 is used as reference model to analyze the influence of 29 input variables on CO<sub>2</sub> emissions forecasting. These input candidates were carefully selected among socioeconomic factors and energy factors that have been considered in the technical literature as major determinant of the CO<sub>2</sub> emission levels.

#### 1) Sensitivity analysis of RBFN parameters

Some observations can be made by analyzing Fig. 3: the RBFN parameters (number of centers and width) greatly affect the forecasting performance of the network. As the number of center increases, the training error decreases. Conversely, the validation error and the testing error showed an opposite behavior. In fact, networks with fewer centers are preferable, since these networks usually have better generalization capabilities, fewer over-fitting problems and they are more

computational efficient. However, if the number of centers is not large enough to capture the underlying behavior of the data, it would result in network with too poor approximation accuracies. The simulation results showed that the optimal number of center is 7.

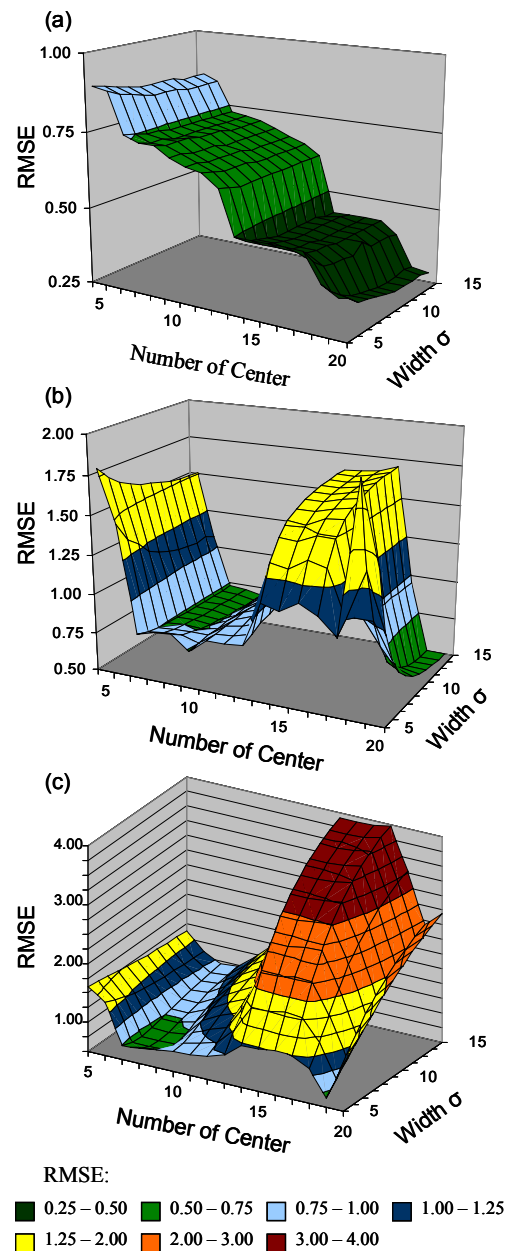


Fig. 3 The effect of RBFN parameters on the CO<sub>2</sub> emissions forecasting performance of the reference model. Data set: (a) training (b) validation (c) testing

In the case of the width parameter, also called as spread parameter, it is important to note that radial basis functions with small width parameter are more selective than functions with large width parameter. In addition, the width parameter should be large enough that the radial basis function for each neuron

responds to overlapping regions of the input space, but not too large that all the neurons respond in essentially the same manner.

In this research,  $\sigma_{\min}$  and  $\sigma_{\max}$  are determined a priori. Besides, the optimal width parameter is obtained by a trial and error procedure. Our analysis showed that for our reference model a width parameter less than 5 will increase the forecasting errors quickly, while any number bigger than 15 does not seem to enhance the forecasting performance or cause it to worsen. The simulation results demonstrated that the optimal width parameter is 8.

## 2) Forecasting Results

The applied RBFN reference model was trained using the data from April 2005 to March 2008. And, the forecasting year is chosen as FY 2008 (April 2008 – March 2009). Fig. 4 illustrates the CO<sub>2</sub> emissions forecasting of Tokyo Electric Power Company. It can be seen from Fig. 4 that the model performs well for the training and prediction periods. Besides, the forecasting accuracy of the reference model is presented in Table VII in terms of RMSE, MAE and MAPE.

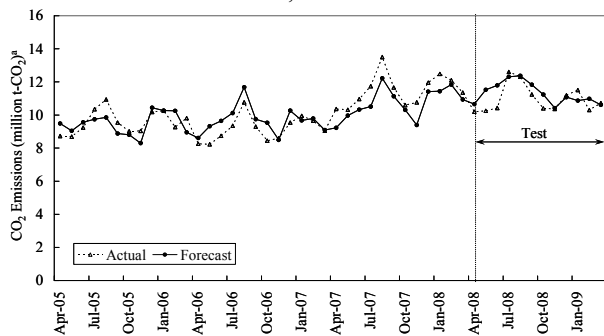


Fig. 4 Results of the CO<sub>2</sub> emissions forecasting using RBFN reference model.

<sup>a</sup>These values don't reflect the use of carbon credits

Data set	RMSE	MAE	MAPE (%)
Training	0.736	0.632	6.51
Validation	0.742	0.636	5.50
Testing	0.691	0.539	4.92

## B. Improvement Rate Results

In this research, the performance of each input candidate is assessed by removing the other inputs from its group and calculating RMSE for this new network under evaluation. Finally, the IR is used to measure the relative improvement between the different competing models and our reference model. When IR is positive, it is mean that the model under evaluation is better than the reference model. The Table VIII presents the results of RMSE calculation for the testing of 29 input candidates. The simulation results showed that the most important input variables are: coal price (fuel price group), coal purchased (fuel use group), trading volume (electric power trading group), thermal generation (electricity data) and Saitama population (population group). Since real GDP is the only input vector inside of the economic indicator group, it was

not remove during the simulation process. In addition, it is observed from Fig. 5 that the population group and the electricity data group had the worst results for IR. This can be attributed to the fact that some inputs inside of these two groups present negative correlation between them (see Table III and Table IV). Therefore, one input variable may not represent the total characteristics of the group. In order to avoid this problem and investigate the forecasting performance of the negative correlated variables, the simulations were running once more for the electricity data group and population group and the results are shown in Table IX. The simulation results demonstrated that two variables (thermal generation and nuclear generation) inside of the electricity data group considerably improved the forecasting performance compared to the previous results.

TABLE VIII  
SIMULATION RESULTS OF THE INPUT VARIABLE TESTING

Variable within the group	Input Variables in model	Number of Center	Width $\sigma$	RMSE
<b>Fuel price group</b>				
Coal price	28	7	12	0.62
Oil price	28	7	11	0.63
LNG price	28	7	8	0.69
<b>Fuel use group</b>				
Coal purchased	23	10	6	0.68
Crude oil purchased	23	10	6	0.69
Heavy oil consumed	23	10	6	0.70
Coal consumed	23	10	9	0.71
Crude oil consumed	23	8	6	0.71
Heavy oil purchased	23	8	6	0.71
LNG consumed	23	8	6	0.73
LNG purchased	23	8	5	0.76
<b>Electric power trading group</b>				
Trading volume	27	6	5	0.80
Day time price	27	6	6	0.82
System peak price	27	6	6	0.82
24-hour mean value price	27	6	6	0.83
<b>Electricity data group</b>				
Thermal generation	24	15	6	0.82
Peak load	24	15	5	0.86
Electricity supply	24	16	5	0.89
Electricity purchased	24	15	5	0.92
Electricity demand	24	16	5	0.94
Hydro generation	24	20	5	0.96
Nuclear generation	24	20	5	0.99
<b>Population group</b>				
Saitama population	25	10	5	0.91
Japan population	25	14	15	0.92
Yokohama population	25	10	5	1.00
Ku-area of Tokyo	25	10	5	1.00
Chiba population	25	10	5	1.08
Kawasaki population	25	10	5	1.10
<b>Economic indicator group</b>				
Real GDP	30	7	8	0.69

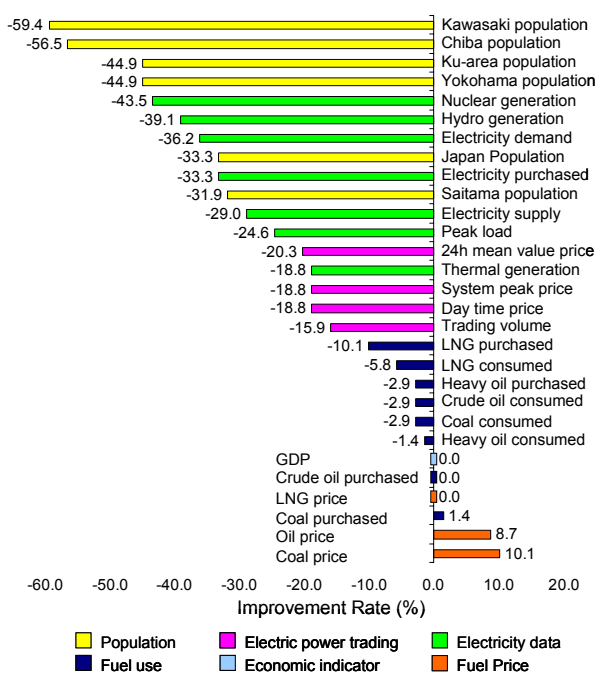


Fig. 5 Improvement rate results of 29 input candidates.

TABLE IX  
RBFN PERFORMANCE UNDER 2 INPUT VARIABLES INSIDE OF POPULATION GROUP AND ELECTRICITY DATA GROUP

Variable within the group	Input Variables in model	Number of Center	Width $\sigma$	RMSE	IR (%)
<b>Electricity data</b>					
E2 and E3	25	5	13	0.72	-4.3
E3 and E5	25	12	6	0.95	-37.7
E1 and E2	25	20	5	0.96	-39.1
E3 and E4	25	5	9	1.05	-52.2
E1 and E4	25	6	10	1.06	-53.6
<b>Population</b>					
P1 and P2	26	14	12	0.80	-15.9
P1 and P5	26	14	10	0.84	-21.7
P1 and P3	26	14	7	0.87	-26.1
P1 and P4	26	14	11	0.88	-27.5
P1 and P6	26	14	10	0.90	-30.4

C. Comparative Analysis

Our input selection procedure identified 8 input variables as the most significant. Namely, coal price, coal purchased by the electric power company, trading volume of electricity, thermal generation, nuclear generation, GDP, Saitama population and the seasonal parameter (month of year). A new RBF network was implemented with the 8 selected input variables and the network parameters were re-calculated. The simulation results showed that the most favorable number of center and width of the new network were 5 and 15 respectively. Fig. 6 presents a comparison of the forecasting results between the reference

model and the new model. Furthermore, the forecasting accuracy measure in terms of RMSE for the reference model and new model were 0.69 and 0.68 respectively. This implies that the new network with 8 input variables is capable of predicting the monthly CO<sub>2</sub> emissions of the electric power company as accurate as the reference model with 30 input variables.

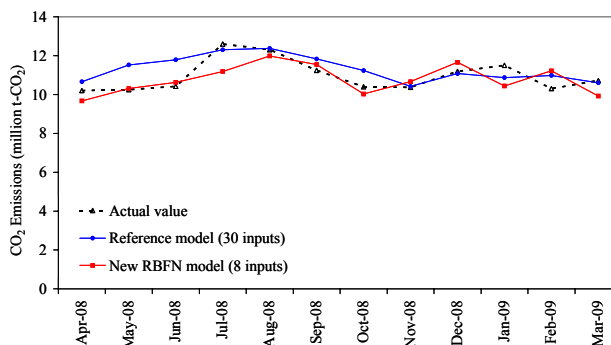


Fig. 6 Comparison of forecasting results between the reference model and the new model.

V. CONCLUSION

This paper introduced an approach for input variable selection for the CO<sub>2</sub> emissions forecasting of an electric power company which is based on identifying the general relationships between groups of input candidates and the output. Eight input variables were identified as the most relevant, which is significantly less than our reference model with 30 input variables.

The results obtained by the new RBFN model are comparable with the reference model results. This lead to the conclusion that by investigating the relationships between the variables, a parsimonious set of input variables can be identified without sacrificing accuracy of the forecasts. Besides, smaller number of input variables leads toward to a more compact neural network that needs less training data and is easier to train.

Moreover, if correlated or insignificant variables are selected as inputs, then computational resources are wasted during ANN training, resulting in prolonged training times and sometimes inferior results.

The number of inputs, the number of centers and width parameter affect the forecasting performance and hence need to be chosen carefully.

REFERENCES

- [1] M. Kainuma, Y. Matsuoka, and T. Morita, "The AIM/end-use model and its application to forecast Japanese carbon dioxide emissions," European Journal of Operational Research, Vol. 122, Issue 2, 16 April 2000, pp. 416-425.
- [2] S. Kasahara, S. Paltsev, J. Reilly, H. Jacoby, and A. D. Ellerman, "Climate Change Taxes and Energy Efficiency in Japan," Journal of Environmental and Resource Economics, Springer, Vol. 37, Number 2, June 2007, pp.377-410.
- [3] C. Yang and S. Schneider, "Global carbon dioxide emissions scenarios: sensitivity to social and technological factors in three regions," Mitigation

- and Adaptation Strategies for Global Change, Springer, Vol. 2, Number 4, 1998, pp. 373-404.
- [4] R. Schmalensee, T. M. Stoker, and R. Judson, "World carbon dioxide emissions: 1950-2050," *Review of Economics and Statistics*, Vol. 80, issue 1, 1998, pp. 15-27.
- [5] D. Holtz-Eakin, and T. M. Selden, "Stoking the fires? CO<sub>2</sub> emissions and economic growth," *Journal of Public Economics*, Vol. 57, Issue 1, May 1995, pp. 85-101.
- [6] A. Islam, and D. Wang, "Forecasting an Electric Utility's CO<sub>2</sub> Emissions Using SAS/AF and SAS/STAT Software: A Linear Analysis," SAS SUGI22 conference proceedings, San Diego - California, March 16-17, 1997.
- [7] A. Lendasse, J. Lee, E. de Bodt, V. Wertz, and M. Verleysen, "Approximation by radial-basis function networks: application to option pricing", ACSEG 2002 conference proceedings Connectionist Approaches in Economics and Management Sciences, Boulogne-sur-Mer, France, 2002, pp. 201-212.
- [8] D. Wedding, K. Cios, "Time series forecasting by combining RBF networks, certainty factors and Box-Jenkins model", *Neurocomputing*, Vol. 10, Issue 2, 1996, pp. 149-168.
- [9] B. Whitehead, T. Choate, "Cooperative-competitive genetic evolution of radial basis function centers and widths for time series prediction", *IEEE Transactions on Neural Networks*, Vol. 7, 1996, pp. 869-880.
- [10] M. J. Orr, "Introduction to Radial Basis Functions Networks," Technical Report, Edinburgh University, Edinburgh, Scotland, UK, April 1996.