



# Forecasting energy demand using neural-network-based grey residual modification models

Yi-Chung Hu<sup>1,2\*</sup> and Peng Jiang<sup>2</sup>

<sup>1</sup>College of Management & College of Tourism, Fujian Agriculture and Forestry University, Fuzhou City, China; and <sup>2</sup>Department of Business Administration, Chung Yuan Christian University, Chung Li Dist., Taoyuan City, Taiwan

Grey forecasting models have taken an important role for forecasting energy demand, particularly the GM(1,1) model, because they are able to construct a forecasting model using a limited samples without statistical assumptions. To improve prediction accuracy of a GM(1,1) model, its predicted values are often adjusted by establishing a residual GM(1,1) model, which together form a grey residual modification model. Two main issues should be considered: the sign estimation for a predicted residual and the way the two models are constructed. Previous studies have concentrated on the former issue. However, since both models are usually established in the traditional manner, which is dependent on a specific parameter that is not easily determined, this paper focuses on the latter issue, incorporating the neural-network-based GM(1,1) model into a residual modification model to resolve the drawback. Prediction accuracies of the proposed neural-network-based prediction models were verified using real power and energy demand cases. Experimental results verify that the proposed prediction models perform well in comparison with original ones.

*Journal of the Operational Research Society* (2017) **68**(5), 556–565. doi:10.1057/s41274-016-0130-2; published online 16 November 2016

**Keywords:** energy demand; forecasting; grey theory; neural network; residual model

## 1. Introduction

Rapid economic development and ongoing industrialization has led to enormous energy consumption. Energy demand prediction has become increasingly important when devising development plans for a country, and particularly for developing countries (Pi *et al.*, 2010), such as China. Meanwhile, energy management is an important issue for future economic prosperity and environmental security (Suganthi and Samuel, 2012). China has become increasingly influential in energy production and consumption (Liu, 2015). China energy has been mainly provided by coal and crude oil, comprising 2/3 (66 %) and 18 %, respectively, of total energy consumed (National Bureau of Statistics of China, 2014). In the past decade, annual energy consumption increases were larger than energy production increases, and China faces not only inevitable environmental impacts, but also the challenge of devising an energy policy forecasting energy demand. Also, energy demand forecasting for China has been an interesting issue (Suganthi and Samuel, 2012).

Grey prediction models are capable of characterizing an unknown system using small data sets (Deng, 1982),

without requiring conformance with statistical assumptions. Only few sample data points are required to achieve reliable and acceptable prediction accuracy (Wen, 2004; Wang and Hsu, 2008) and have been widely applied to management, economics, and engineering (e.g. Feng *et al.*, 2012; Chang *et al.*, 2015; Pi *et al.*, 2010; Lee and Tong, 2011; Mao and Chirwa, 2006; Zeng *et al.*, 2016; Chang *et al.*, 2015; Tsaur and Liao, 2007; Wen, 2004; Cui *et al.*, 2013; Wei *et al.*, 2015). Grey prediction systems are particularly appropriate for energy demand forecasting, because energy consumption data are often few and/or do not conform to the usual statistical assumptions, such as normal distribution (Lee and Tong, 2011; Suganthi and Samuel, 2012). Wu *et al.*, (2013) also used practical numerical examples to demonstrate that the small sample usually has more accuracy than the large sample for setting up a grey prediction model. Artificial intelligence techniques, multivariate regression, and time series models (e.g. Ediger and Akar, 2007; Gonzalez and Zamarreno, 2005; Tutun *et al.*, 2015; Lauret *et al.*, 2008; Duran, 2009; Xia *et al.*, 2010) require large sample sizes to achieve reasonable forecasting accuracy (Wang and Hsu, 2008; Feng *et al.*, 2012; Chang *et al.*, 2015; Pi *et al.*, 2010), which is impractical for energy demand forecasting.

The GM(1,1) model is one of the most frequently used grey prediction models for time series forecasting (Liu and Lin,

\*Correspondence: Yi-Chung Hu, Department of Business Administration, Chung Yuan Christian University, Chung Li Dist., Taoyuan City, Taiwan. E-mail: ychu@cycu.edu.tw

2006). To improve prediction accuracy of the original GM(1,1) model, several improved versions have been proposed, such as a discrete forecasting model (Xie and Liu, 2009), a grey Lotka–Volterra model (Wu *et al.*, 2012), a new model with the fractional order accumulation (Wu *et al.*, 2013), a novel BGM(1,1) using a box plot to analyse data features (Chang *et al.*, 2015), an improved grey model with convolution integral GMC(1,  $n$ ) (Wang and Hao, 2016), and a self-adaptive intelligence model (Zeng *et al.*, 2016). Besides, the residual model has been eye-catching and played an important role in grey prediction (Liu and Lin, 2006; Deng, 1982). The residual modification model thus becomes the focus of this study.

When the corresponding residual model is established, predicted values from the original model can be adjusted by those from the residual model. The two models comprise a grey residual modification model, and both models are usually constructed in the same way as the traditional GM(1,1) model. However, the background value has an important role for the traditional model, but is not easily determined. This leads to some well-known prediction models using the traditional GM(1,1) model, developed for residual sign estimation to improve the prediction accuracy of the residual modification model, which also encounter the difficulty in determining background value, for example the MLP-GM(1,1) model based on multi-layer perceptron (MLP) (Hsu and Chen, 2003) and the GP-GM(1,1) model based on genetic programming (GP; Lee and Tong, 2011). The neural-network-based GM(1,1) (NN-GM(1,1)) model is free of the dependency on the background value (Hu *et al.*, 2001) and performs well in comparison with the traditional GM(1,1) model. Therefore, it is interesting to investigate the impact on prediction accuracy of the proposed neural-network-based residual modification models on energy demand forecasting for China by using the NN-GM(1,1) model rather than the traditional GM(1,1) model.

The remainder of the paper is organized as follows. Section 2 introduces the traditional grey residual modification model, and Section 3 introduces the NN-GM(1,1) model and the proposed neural-network-based residual modification model. On the basis of the MLP-GM(1,1) and GP-GM(1,1) models, Section 4 examines the forecasting performances of the proposed prediction models using real cases of power and energy demand. Section 5 discusses the outcomes and presents conclusions.

## 2. Traditional grey residual modification model

### 2.1. Traditional GM(1,1) model

The computational steps to construct a traditional GM(1,1) model are as follows:

Step 1 Present an original and nonnegative data sequence  $\mathbf{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$ , provided by one system and consisting of  $n$  samples.

Step 2 Perform the accumulated generating operation (AGO).

Identify the potential regularity hidden in  $\mathbf{x}^{(0)}$  using AGO (Liu and Lin, 2006; Duran, 2009) to generate a new sequence,  $\mathbf{x}^{(1)} = (x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)})$ ,

$$x_k^{(1)} = \sum_{j=1}^k x_j^{(0)}, \quad k = 1, 2, \dots, n \quad (1)$$

and  $x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}$  can be then approximated by a first-order differential equation,

$$\frac{d\mathbf{x}^{(1)}}{dt} + a\mathbf{x}^{(1)} = b \quad (2)$$

where  $a$  and  $b$  are the developing coefficient and control variable, respectively. The predicted value,  $\hat{x}_k^{(1)}$ , for  $x_k^{(1)}$  can be obtained by solving the differential equation with initial condition  $x_1^{(1)} = x_1^{(0)}$ :

$$\hat{x}_k^{(1)} = \left(x_1^{(0)} - \frac{b}{a}\right)e^{-a(k-1)} + \frac{b}{a} \quad (3)$$

Step 3 Determine the developing coefficient and control variable.

$a$  and  $b$  can be obtained using the ordinary least-squares method:

$$[a, b]^T = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{y} \quad (4)$$

where

$$\mathbf{B} = \begin{bmatrix} -z_2^{(1)} & 1 \\ -z_3^{(1)} & 1 \\ \vdots & \vdots \\ -z_n^{(1)} & 1 \end{bmatrix} \quad (5)$$

$$z_k^{(1)} = \alpha x_k^{(1)} + (1 - \alpha)x_k^{(1)} \quad (6)$$

$$\mathbf{y} = [x_2^{(0)}, x_3^{(0)}, \dots, x_n^{(0)}]^T \quad (7)$$

where  $z_k^{(1)}$  is the background value.  $\alpha$  is usually specified as 0.5 for convenience, but this is not an optimal setting. Thus,  $a$  and  $b$  are fully dependent on  $z_k^{(1)}$ , which is not easily determined.

Step 4 Perform the inverse accumulated generating operation (IAGO).

Using the IAGO, the predicted value of  $x_k^{(0)}$  is

$$\hat{x}_k^{(0)} = \hat{x}_k^{(1)} - \hat{x}_{k-1}^{(1)}, \quad k = 2, 3, \dots, n \quad (8)$$

Therefore,

$$\hat{x}_k^{(0)} = (1 - e^a) \left( x_1^{(0)} - \frac{b}{a} \right) e^{-a(k-1)}, \quad k = 2, 3, \dots, n \quad (9)$$

and note that  $\hat{x}_1^{(1)} = \hat{x}_1^{(0)}$  holds.

### 2.2. Residual modification using traditional GM(1,1) models

A residual modification model is usually established using traditional GM(1,1) models. The computational steps of constructing a traditional residual modification model are as follows:

Step 1 Establish a traditional GM(1,1) model for  $x^{(0)}$ .

Step 2 Generate the sequence of absolute residual values,  $\epsilon^{(0)} = (\epsilon_2^{(0)}, \epsilon_3^{(0)}, \dots, \epsilon_n^{(0)})$ , where

$$\epsilon_k^{(0)} = |x_k^{(0)} - \hat{x}_k^{(0)}|, \quad k = 2, 3, \dots, n \quad (10)$$

Step 3 Establish a residual model.

A residual model is established as a traditional GM(1,1) model for  $\epsilon^{(0)}$ . Similar to  $\hat{x}_k^{(0)}$ , the predicted residual of  $\epsilon_k^{(0)}$  is

$$\hat{\epsilon}_k^{(0)} = (1 - e^{a_e}) \left( \epsilon_2^{(0)} - \frac{b_e}{a_e} \right) e^{-a_e(k-1)}, \quad k = 3, 4, \dots, n \quad (11)$$

where  $a_e$  and  $b_e$  are the developing coefficient and the control variable, respectively, and are also fully dependent on the background value.

Step 4 Perform residual modification.

A predicted value  $\hat{x}_{k^{ir}}^{(0)}$  can be obtained by adding or subtracting  $\hat{\epsilon}_k^{(0)}$  from original  $\hat{x}_k^{(0)}$  (Hsu and Wen, 1998).

$$\hat{x}_{k^{ir}}^{(0)} = \hat{x}_k^{(0)} + s_k \hat{\epsilon}_k^{(0)}, \quad k = 2, 3, \dots, n \quad (12)$$

where  $s_k$  denotes the positive or negative sign for  $\hat{\epsilon}_k^{(0)}$ . The determination of  $s_k$  can be dependent on the mechanism of sign estimation provided by other residual modification models, for instance the MLP-GM(1,1) and GP-GM(1,1) models. For simplicity, the sign estimation methods of those two prediction models are omitted.

## 3. Neural-network-based grey residual modification model

### 3.1. NN-GM(1,1) model

Because  $z_k^{(1)}$  is not easily determined, it is quite reasonable to consider finding  $a$  and  $b$  without requiring  $z_k^{(1)}$ . A cost function  $E(a, b)$ ,

$$E(a, b) = \frac{1}{2} \sum_k \left( x_k^{(0)} - \hat{x}_k^{(0)} \right)^2, \quad k = 2, 3, \dots, n \quad (13)$$

was built for the NN-GM(1,1) model, where  $a$  and  $b$  are the connection weights. The model itself was a widely used single-layer perceptron (SLP). Similar to the back-propagation algorithm (BP; Smith and Gupta, 2002), the computational steps to constructing such a model are as follows:

Step 1 Present a randomly selected sequence  $(k, 1, 1)$  ( $k = 2, 3, \dots, n$ ) with  $x_k^{(0)}$  as its desired output to NN-GM(1,1).

Step 2 Calculate the actual output  $\hat{x}_k^{(0)}$  of NN-GM(1,1).

Step 3 Adjust the connection weights. For  $(k, 1, 1)$ ,  $a$  and  $b$  are adjusted to  $a + \Delta a$  and  $b + \Delta b$ , respectively. Then  $\Delta a$  and  $\Delta b$  can be derived by the gradient descent method on the cost function, and

$$\Delta a = \eta \left( x_k^{(0)} - \hat{x}_k^{(0)} \right) V_{ak} \quad (14)$$

$$\Delta b = \eta \left( x_k^{(0)} - \hat{x}_k^{(0)} \right) V_{bk} \quad (15)$$

where

$$V_{ak} = \left[ (-e^a) \left( x_1^{(0)} - \frac{b}{a} \right) e^{-a(k-1)} + (1 - e^a) \left( \frac{b}{a^2} \right) e^{-a(k-1)} + (1 - e^a) \left( x_1^{(0)} - \frac{b}{a} \right) (-k + 1) e^{-a(k-1)} \right] \quad (16)$$

$$V_{bk} = \left[ (1 - e^a) \left( -\frac{1}{a} \right) e^{-a(k-1)} \right] \quad (17)$$

Step 4 Terminate when a pre-specified number of iterations have been performed; otherwise, return to Step 1.

### 3.2. Residual modification using NN-GM(1,1) models

In the proposed neural-network-based model, traditional GM(1,1) models are no longer used; rather, NN-GM(1,1) models considered. The construction of the proposed grey prediction model is described as follows:

Step 1 Establish a NN-GM(1,1) model for  $x^{(0)}$ .

Step 2 For  $\mathbf{x}^{(0)}$ , generate the sequence of absolute residual values,  $\boldsymbol{\varepsilon}^{(0)} = (\varepsilon_2^{(0)}, \varepsilon_3^{(0)}, \dots, \varepsilon_n^{(0)})$ .

Step 3 Establish a residual NN-GM(1,1) model. A residual NN-GM(1,1) model, using all sequences of absolute residual values, is established where  $a_e$  and  $b_e$  are connection weights of a SLP for processing residuals.  $\Delta a_e$  and  $\Delta b_e$  may be derived with respect to  $a_e$  and  $b_e$ , respectively, by defining a cost function

$$E(a_e, b_e) = \frac{1}{2} \sum_k (\varepsilon_k^{(0)} - \hat{\varepsilon}_k^{(0)})^2, \quad k = 3, 4, \dots, n \tag{18}$$

Similar to  $a$  and  $b$ ,  $a_e$  and  $b_e$  can be adjusted to  $a_e + \Delta a_e$  and  $b_e + \Delta b_e$ , respectively, after presenting a randomly selected sequence  $(k, 1, 1)$  ( $k = 3, 4, \dots, n$ ) with desired output  $\varepsilon_k^{(0)}$  to the SLP related to, where

$$\Delta a_e = \eta (\varepsilon_k^{(0)} - \hat{\varepsilon}_k^{(0)}) V_{ak} \tag{19}$$

$$\Delta b_e = \eta (\varepsilon_k^{(0)} - \hat{\varepsilon}_k^{(0)}) V_{bk} \tag{20}$$

and

$$\begin{aligned} V_{ak} &= (-e^a) \left( \varepsilon_2^{(0)} - \frac{b}{a} \right) e^{-a(k-1)} \\ &\quad + (1-e^a) \left( \frac{b}{a^2} \right) e^{-a(k-1)} \\ &\quad + (1-e^a) \left( \varepsilon_2^{(0)} - \frac{b}{a} \right) (-k+1) e^{-a(k-1)} \end{aligned} \tag{21}$$

$$V_{bk} = (1-e^a) \left( -\frac{1}{a} \right) e^{-a(k-1)} \tag{22}$$

Step 4 Perform residual modification.

A predicted value  $\hat{x}_{k^{nmr}}^{(0)}$  is produced by adding or subtracting  $\hat{\varepsilon}_k^{(0)}$  from the original  $\hat{x}_k^{(0)}$ ,

$$\hat{x}_{k^{nmr}}^{(0)} = \hat{x}_k^{(0)} + s_k \hat{\varepsilon}_k^{(0)}, \quad k = 2, 3, \dots, n \tag{23}$$

As demonstrated in Figure 1, two independent SLPs were employed to establish the proposed neural-network-based grey residual modification model: one each for the original and residual sequences. The MLP-GM(1,1) and GP-GM(1,1) models used two traditional GM(1,1) models, independently, and artificial intelligence tools were applied to effectively determine  $s_k$ . The flow chart of the proposed residual modification model is illustrated in Figure 2.

When the NN-GM(1,1) models were incorporated into the MLP-GM(1,1) and GP-GM(1,1) models, rather traditional

GM(1,1) models, two new prediction models, NN-MLP-GM(1,1) and NN-GP-GM(1,1), were able to remove the requirement of determining background values.

### 4. Empirical results

Empirical studies were conducted using real data sets to compare energy demand forecasting ability of the proposed NN-MLP-GM(1,1) and NN-GP-GM(1,1) models against original GM(1,1), NN-GM(1,1), MLP-GM(1,1), and NN-GP-GM(1,1) models. Mean absolute percentage error (MAPE) was employed to measure prediction performance, as this can be treated as the benchmark and is more stable than the commonly used mean absolute error and root mean square error (Makridakis, 1993; Lee and Shih, 2011). MAPE with respect to  $x_k^{(0)}$  is

$$\text{MAPE} = \sum_{k \in T} \frac{e_k}{|T|} \tag{24}$$

where  $T$  denotes the set of training or test data, whereas  $e_k$  is the absolute percentage error (APE) with respect to  $x_k^{(0)}$ ,

$$e_k = \frac{|x_k^{(0)} - \hat{x}_{k^{pr}}^{(0)}|}{x_k^{(0)}} \times 100\% \tag{25}$$

where  $\hat{x}_{k^{pr}}^{(0)}$  is a predicted value (e.g.  $\hat{x}_k^{(0)}$ ,  $\hat{x}_{k^{tr}}^{(0)}$ ,  $\hat{x}_{k^{nmr}}^{(0)}$ ) with respect to  $x_k^{(0)}$ . Lewis (1982) proposed MAPE criteria for evaluating a forecasting model, where  $\text{MAPE} \leq 10$ ,  $10 < \text{MAPE} \leq 20$ ,  $20 < \text{MAPE} \leq 50$ , and  $\text{MAPE} > 50$  correspond to high, good, reasonable, and weak forecasting models, respectively.

#### 4.1. Applications to energy demand forecasting

4.1.1. Case I An experiment was conducted on the historical annual power demand of Taiwan from 1985 to 2000. As in Hsu and Chen (2003), data from 1985 to 1998 were reserved for the model-fitting, and data from 1999 to 2000 were used for ex post testing. Table 1 summarizes forecasting results, reported by Hsu and Chen (2003), of original GM(1,1) and MLP-GM(1,1) models, along with the corresponding details for the proposed NN-MLP-GM(1,1) model. From Table 1, we can see that the MAPE of the original GM(1,1), the MLP-GM(1,1), and the NN-MLP-GM(1,1) models for model-fitting was 1.54, 0.57, and 1.56 %, respectively. And for ex post testing, the MAPE was 3.88, 1.29, and 0.78 %, respectively.

It is noteworthy that, although the NN-MLP-GM(1,1) model is slightly inferior to the original GM(1,1) and the MLP-GM(1,1) models for model-fitting, it is superior to the original GM(1,1) and the MLP-GM(1,1) models for ex post testing. Actually, when evaluating a prediction model, more emphasis should be placed on generalization rather than model-fitting (Luo et al., 2013). In this case, the MLP-GM(1,1) model seems

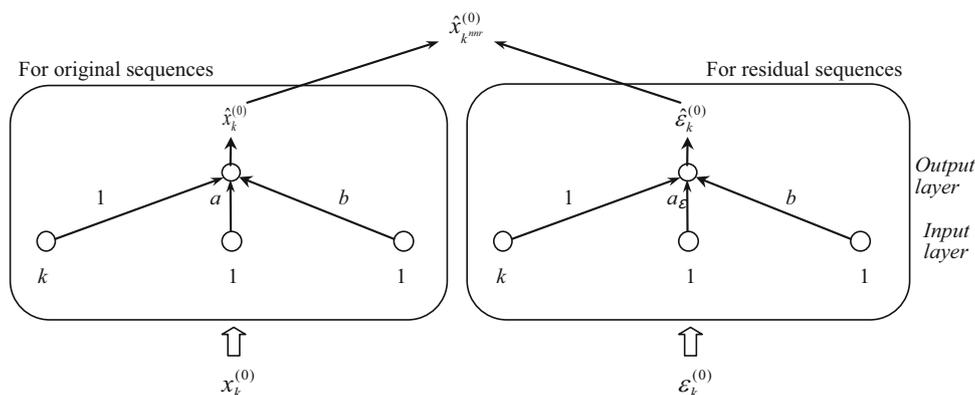


Figure 1 A neural-network-based residual modification model.

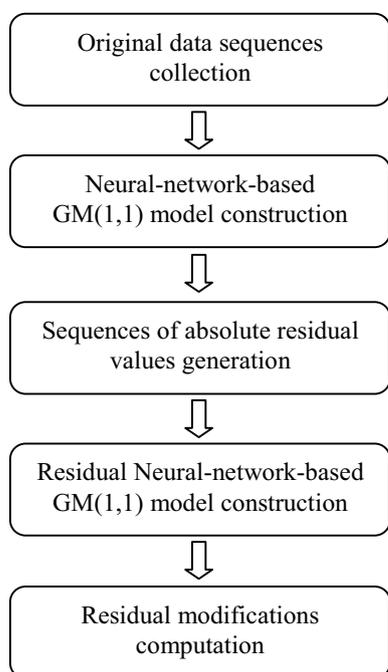


Figure 2 Flow chart of the proposed residual modification model.

to suffer from over-fitting. Figure 3 demonstrates the superiority of the generalization ability of the proposed NN-MLP-GM(1,1) model over the original GM(1,1) and the MLP-GM(1,1) models.

4.1.2. Case II The second experiment was conducted on the historical annual energy demand of China, collected from 1990 to 2007. Same as Lee and Tong (2011), data from 1990 to 2003 were used for the model-fitting, and data from 2004 to 2007 were used for ex post testing. Forecasting results from Lee and Tong (2011) obtained by original GM(1,1), MLP-GM(1,1) and GP-GM(1,1) models are summarized in Table 2, along with the corresponding details for the proposed NN-

MLP-GM(1,1) and NN-GP-GM(1,1) models. Table 2 shows that the MAPE of the original GM(1,1), the MLP-GM(1,1), the GP-GM(1,1), the NN-GM(1,1), the NN-MLP-GM(1,1), and the NN-GP-GM(1,1) models for model-fitting was 4.13, 3.61, 2.59, 3.81, 4.15, and 2.80 %, respectively. And for ex post testing, the MAPE was 26.21, 20.23, 20.23, 28.71, 14.81, and 14.81 %, respectively. Since a change on an epic scale happened to 2004, this can explain why results of the ex post testing is not as good as those of the model-fitting.

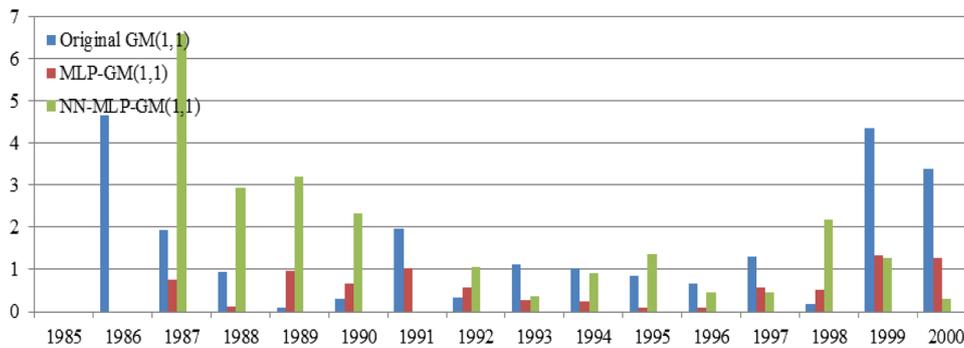
Similar to Case I, although MAPE obtained by NN-MLP-GM(1,1) and NN-GP-GM(1,1) models is slightly inferior to that from MLP-GM(1,1) and GP-GM(1,1), respectively, for model-fitting, they are superior to MLP-GM(1,1) and GP-GM(1,1), respectively, for ex post testing. In this case, it seems that both MLP-GM(1,1) and GP-GM(1,1) models suffer from over-fitting. The predicted values obtained by different forecasting models are illustrated in Figure 4. The proposed NN-MLP-GM(1,1) and NN-GP-GM(1,1) models show “good” generalization ability, whereas the other prediction models have only “reasonable” forecasting ability for testing data. The generalization ability of the NN-MLP-GM(1,1) and the NN-GP-GM(1,1) models are conspicuous.

4.1.3. Case III The third experiment was conducted on historical annual electricity demand of China, collected from China Statistical Yearbook (National Bureau of Statistics of China, 2014), 1981–2002. Following (Zhou et al, 2006), data from 1981 to 1998 were used for model-fitting, and from 1999 to 2002 for ex post testing. The forecasting results obtained from the different forecasting models are summarized in Table 3. All the models have “high” forecasting ability on the training and test data.

Table 3 shows that results obtained by the proposed NN-MLP-GM(1,1) and NN-GP-GM(1,1) models are satisfactory. The MAPE of the original GM(1,1), the MLP-GM(1,1), the GP-GM(1,1), the NN-GM(1,1), the NN-MLP-GM(1,1), and the NN-GP-GM(1,1) models for model-fitting was 2.28, 2.03, 1.44, 1.84, 1.84, and 1.28 %, respectively. And for ex post

**Table 1** Prediction accuracy obtained by different forecasting models for power demand (unit: 10<sup>3</sup> Wh)

Year	Actual	Original GM(1,1)		MLP-GM(1,1)		NN-MLP-GM(1,1)	
		Predicted	APE	Predicted	APE	Predicted	APE
1985	47919102	47919102	0	47919102	0	47919102	0
1986	53812862	56318092	4.66	53812862	0	53812862	0
1987	59174751	60319829	1.94	59630904	0.77	63072052.4	6.59
1988	65227727	64605914	0.95	65310510	0.13	67141427.8	2.93
1989	69251809	69196550	0.08	69917174	0.96	71470372.97	3.20
1990	74344947	74113379	0.31	74850394	0.68	76076268.3	2.33
1991	80977405	79379577	1.97	80133358	1.04	80977428.0	0
1992	85290354	85019971	0.32	85790897	0.59	86193211.8	1.059
1993	92084684	91061148	1.11	91849611	0.26	91751490.1	0.36
1994	98561004	97531587	1.04	98337985	0.23	97657263.9	0.92
1995	105368193	104461790	0.86	105286530	0.08	103943602.7	1.35
1996	111139816	111884424	0.67	111040924	0.09	110634911.7	0.45
1997	118299046	119834482	1.30	118971794	0.57	117757193.7	0.46
1998	128129801	128349438	0.17	127467127	0.52	125338144.8	2.18
MAPE			1.54		0.57		1.56
1999	131725892	137469433	4.36	133459644	1.32	133406202.6	1.28
2000	142412887	147237458	3.39	144204700	1.26	141995935.9	0.29
MAPE			3.88		1.29		0.78



**Figure 3** Absolute percentage errors by different prediction models for Case I.

testing, the MAPE was 7.24, 3.90, 3.90, 10.35, 3.34, and 3.34 %, respectively. The proposed NN-MLP-GM(1,1) and NN-GP-GM(1,1) models have superior fitting and generalization ability compared to MLP-GM(1,1) and GP-GM(1,1) models, respectively. Figure 5 also demonstrates the superiority of the generalization ability of NN-MLP-GM(1,1) and NN-GP-GM(1,1) models over the other prediction models.

**5. Discussion and conclusions**

Energy demand forecasting can be regarded as a grey system problem (Pi *et al.*, 2010; Suganthi and Samuel, 2012) because several factors, such as income and population, influence energy demand but the precise relationships are not clear. That is, although relationships exist between input factors and dependent variable in the real problems, but it is not distinct about what these relationships are (Hu, 2016; Hu *et al.*, 2015).

Energy demand data are often limited and do not conform to the usual statistical assumptions, such as normal distribution. The GM(1,1) model is the most frequently used grey prediction model and has played an important role in energy demand prediction because it requires only limited samples to construct a prediction model without statistical assumptions. However, the traditional residual modification model has suffered from determination of the background value, as does the traditional GM(1,1) model, whereas the NN-GM(1,1) model is able to directly determine the developing coefficient and control variable using a SLP without requiring the background value. The NN-GM(1,1) model is also simple to implement as a computer program. Therefore, it is reasonable to replace the traditional GM(1,1) model with the NN-GM(1,1) model for a grey residual modification model. It is noted that, unlike the traditional SLP, the NN-GM(1,1) model does not use the sigmoid function as its activation function.

**Table 2** Prediction accuracy obtained by different forecasting models for energy demand (unit: 10<sup>4</sup> tons of SCE)

Year	Actual	Original GM(I,I)		MLP-GM(I,I)		GP-GM(I,I)		NN-GM(I,I)		NN-MLP-GM(I,I)		NN-GP-GM(I,I)	
		Predicted	APE	Predicted	APE	Predicted	APE	Predicted	APE	Predicted	APE	Predicted	APE
1990	98703	98703	0	98703	0	98703	0	98703	0	98703	0	98703	0
1991	103783	108706.1	4.74	103783	0	103783	0	108144	4.20	103783	0	103783	0
1992	109170	112335.5	2.9	116225.8	6.46	108445.2	0.66	111529	2.16	114105.3	4.52	108952.25	0.20
1993	115993	116086.1	0.08	111804.1	3.61	111804.1	3.61	115019	0.84	116830.0	0.72	116830.03	0.72
1994	122737	119962	2.26	115248.8	6.10	124675.1	1.58	118619	3.35	120901.3	1.50	120901.27	1.50
1995	131176	123967.2	5.50	129154.8	1.54	129154.8	1.54	122332	6.74	125207.9	4.55	125207.89	4.55
1996	138948	128106.2	7.80	133816.1	3.69	133816.1	3.69	126160	9.20	129785.4	6.59	129785.38	6.59
1997	137798	132383.3	3.93	138668.2	0.63	138668.2	0.63	130109	5.58	134677.6	2.26	134677.63	2.26
1998	132214	136803.3	3.47	143721	8.7	129885.5	1.76	134181	1.49	139939.1	5.84	128423.23	2.87
1999	133831	141370.8	5.63	133756.5	0.06	133756.5	0.06	138381	3.40	131123.8	2.02	131123.8	2.02
2000	138553	146090.8	5.44	137709.8	0.61	137709.8	0.61	142712	3.00	151858.0	9.60	133565.56	3.60
2001	143199	150968.4	5.43	141743.6	1.02	141743.6	1.02	147178	2.78	158705.7	10.83	135651.02	5.27
2002	151797	156008.9	2.77	145855.2	3.91	145855.2	3.91	151785	0.01	166313.1	9.56	166313.08	9.56
2003	174990	161217.6	7.87	150041.6	14.26	172393.5	1.48	156535	10.55	174845.9	0.08	174845.89	0.08
MAPE			4.13		3.61		2.59		3.81		4.15		2.80
2004	203227	166600.2	18.02	178901.5	11.97	178901.5	11.97	161434	20.56	184512.1	9.21	184512.1	9.21
2005	224682	172162.6	23.37	185702.4	17.35	185702.4	17.35	166487	25.90	195572.6	12.96	195572.6	12.96
2006	264270	177910.7	32.68	192813.8	27.04	192813.8	27.04	171698	35.03	208355.3	21.16	208355.3	21.16
2007	265583	183850.7	30.77	200254.3	24.60	200254.3	24.60	177072	33.33	223272.4	15.93	223272.4	15.93
MAPE			26.21		20.23		20.23		28.71		14.81		14.81

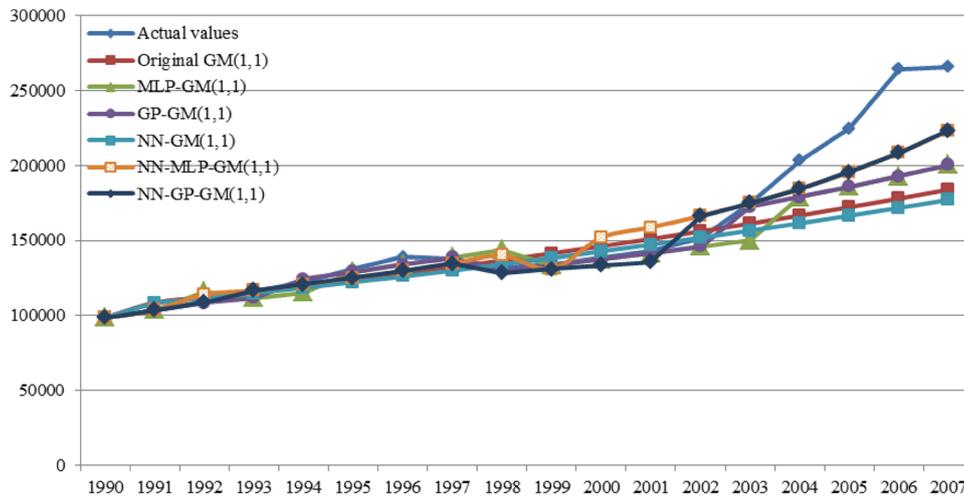


Figure 4 Predicted values obtained by different prediction models.

Table 3 Prediction accuracy obtained by different forecasting models for electricity demand (unit: 100 million kWh)

Year	Actual	Original GM(1,1)		MLP-GM(1,1)		GP-GM(1,1)		NN-GM(1,1)		NN-MLP-GM(1,1)		NN-GP-GM(1,1)	
		Predicted	APE	Predicted	APE	Predicted	APE	Predicted	APE	Predicted	APE	Predicted	APE
1981	3096	3096	0	3096	0	3096	0	3096	0	3096	0	3096	0
1982	3280	3327.7	1.45	3280	0	3280	0	3230.1	1.52	3280	0	3280	0
1983	3519	3611.5	2.63	3638.0	3.38	3585.0	1.88	3516.7	0.07	3551.4	0.92	3551.4	0.92
1984	3778	3919.5	3.75	3888.3	2.92	3888.3	2.92	3828.7	1.34	3804.6	0.70	3804.6	0.70
1985	4118	4253.9	3.30	4217.1	2.41	4217.1	2.41	4168.4	1.22	4138.1	0.49	4138.1	0.49
1986	4507	4616.7	2.43	4573.3	1.47	4573.3	1.47	4538.3	0.69	4500.3	0.15	4500.3	0.15
1987	4985	5010.5	0.51	4959.4	0.51	4959.4	0.51	4940.9	0.88	4893.4	1.84	4988.5	0.07
1988	5467	5437.9	0.53	5377.6	1.63	5498.2	0.57	5379.3	1.60	5438.9	0.51	5438.9	0.51
1989	5865	5901.7	0.63	5830.7	0.59	5830.7	0.59	5856.5	0.14	5931.2	1.13	5931.2	1.13
1990	6230	6405.1	2.81	6321.4	1.47	6321.4	1.47	6376.2	2.35	6469.7	3.85	6469.7	3.85
1991	6775	6951.4	2.60	6852.8	1.15	6852.8	1.15	6941.9	2.46	7059.1	4.19	6824.7	0.73
1992	7542	7544.3	0.03	7428.1	1.51	7428.1	1.51	7557.8	0.21	7411.0	1.74	7411.0	1.74
1993	8426.5	8187.8	2.83	8050.8	4.46	8324.8	1.21	8228.3	2.35	8044.4	4.53	8044.4	4.53
1994	9260.4	8886.2	4.04	8724.8	5.78	9047.6	2.30	8958.4	3.26	9188.9	0.77	9188.9	0.77
1995	10023.4	9644.1	3.78	9834.3	1.89	9834.3	1.89	9753.2	2.70	10042	0.19	10042	0.19
1996	10764.3	10466.7	2.76	10690.9	0.68	10690.9	0.68	10618.6	1.35	10980.5	2.01	10980.5	2.01
1997	11284.4	11359.5	0.67	11623.7	3.01	11095.3	1.68	11560.7	2.45	12014.1	6.47	11107.2	1.57
1998	11598.4	12328.4	6.29	12017.1	3.61	12017.1	3.61	12586.4	8.52	12018.2	3.62	12018.2	3.62
MAPE			2.28		2.03		1.44		1.84		1.84		1.28
1999	12305.2	13379.9	8.73	13013	5.75	13013.0	5.75	13703.1	11.36	12991.2	5.57	12991.2	5.57
2000	13471.4	14521.2	7.79	14088.8	4.58	14088.8	4.58	14918.9	10.74	14026.9	4.12	14026.9	4.12
2001	14633.5	15759.8	7.70	15250.3	4.21	15250.3	4.21	16242.5	10.99	15124.9	3.36	15124.9	3.36
2002	16331.5	17104	4.73	16503.6	1.05	16503.5	1.05	17683.6	8.279	16283.3	0.30	16283.3	0.30
MAPE			7.24		3.90		3.90		10.35		3.34		3.34

Some improved residual modification models, such as MLP-GM(1,1) and GP-GM(1,1), focused on residual sign estimation, but they retain the drawback of the traditional GM(1,1) model. On the other hand, the proposed NN-MLP-GM(1,1) and NN-GP-GM(1,1) models were developed from MLP-GM(1,1) and GP-GM(1,1) models, respectively, by substituting NN-GM(1,1) for traditional GM(1,1) models. Therefore, the proposed residual modification model can

estimate residual signs effectively and is free from the drawback of the traditional GM(1,1) model.

Real cases of energy demand data from China were used to evaluate the forecasting performances of the proposed NN-MLP-GM(1,1) and NN-GP-GM(1,1) models. The outcomes verified that the proposed forecasting models perform well. Zhou *et al* (2006) showed that for Case III, MAPE for an autoregressive integrated moving average (ARIMA) and

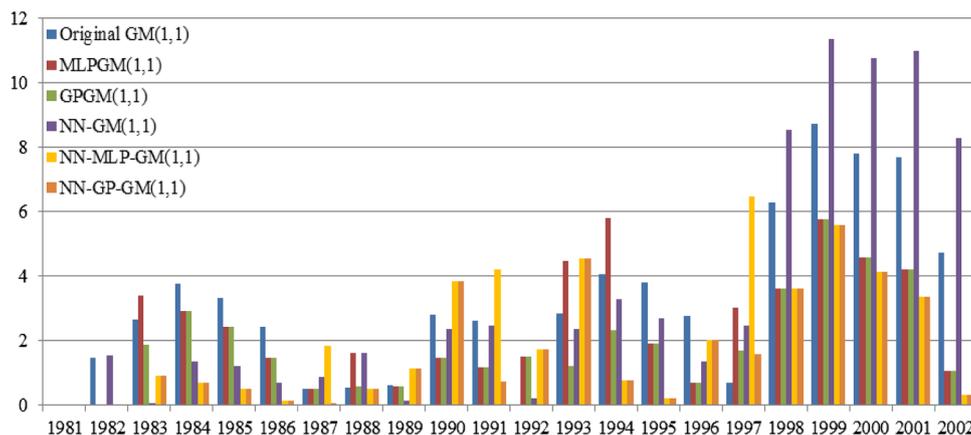


Figure 5 Absolute percentage errors by different prediction models for Case III.

trigonometric grey prediction model was 3.25 and 2.12 %, for model-fitting, respectively, which are inferior to the proposed NN-MLP-GM(1,1) and NN-GP-GM(1,1) models. For Cases II and III, it is interesting to note that the NN-GM(1,1) model is superior to the traditional GM(1,1) model for model-fitting, but inferior for ex post testing. In other words, the NN-GM(1,1) model appears to be over-fitting. Experimental results show that the generalization ability of NN-MLP-GM(1,1) and NN-GP-GM(1,1) models are superior to the MLP-GM(1,1) and GP-GM(1,1) models. Thus, the generalization ability of a residual modification model could be improved by incorporating NN-GM(1,1) models.

The SLP in this study was trained on the basis of the BP using gradient descent. The learning is continued until a convergent condition is reached. It is known that one drawback of using BP is that a local minimum (Weiss and Kulikowski, 1991) is likely to be stuck during the learning process. Therefore, other optimization techniques such as genetic algorithm (GA; Goldberg, 1989; Man *et al.*, 1999) could be applied to automatically determine the connection weights. In parenthesis, in comparison with the BP, an advantage of using GA is that a local minimum is unlikely to be stuck (Rooij *et al.*, 1996; Vonkij *et al.*, 1997; Hu, 2010). Additionally, the MLP-GM(1,1), the NN-GP-GM(1,1), and the proposed models have something in common. It is evident that they are grey residual modification models and developed for residual sign estimation to improve the prediction accuracy of the residual modification model. However, it is interesting to estimate not only the sign but the extent to which  $\hat{x}_k^{(0)}$  obtained from the original GM(1,1) model can be modified by  $\hat{\varepsilon}_k^{(0)}$  ( $k = 2, 3, \dots, n$ ). This remains for the future work.

*Acknowledgments*—The authors would like to thank the anonymous referees for their valuable comments. This research is partially supported by the Ministry of Science and Technology, Taiwan, under Grant MOST 104-2410-H-033-023-MY2.

## References

- Chang CJ, Dai WL and Chen CC (2015). A novel procedure for multimodel development using the grey silhouette coefficient for small-data-set forecasting. *Journal of the Operational Research Society* **66**(11):1887–1894.
- Chang CJ, Li DC, Huang YH and Chen CC (2015). A novel gray forecasting model based on the box plot for small manufacturing data sets. *Applied Mathematics and Computation* **265**(C):400–408.
- Cui J, Liu SF, Zeng B and Xie NM (2013). A novel grey forecasting model and its optimization. *Applied Mathematical Modelling* **37**(6):4399–4406.
- Deng JL (1982). Control problems of grey systems. *Systems and Control Letters* **1**(5):288–294.
- Duran Toksari M (2009). Estimating the net electricity energy generation and demand using ant colony optimization approach: Case of Turkey. *Energy Policy* **37**(3):1181–1187.
- Ediger VS and Akar S (2007). ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy* **35**(3):1701–1708.
- Feng SJ, Ma YD, Song ZL and Ying J (2012). Forecasting the energy consumption of China by the grey prediction model. *Energy Sources, Part B: Economics, Planning, and Policy* **7**(4):376–389.
- Goldberg DE (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. MA: Addison-Wesley.
- Gonzalez PA and Zamarreno JA (2005). Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and Buildings* **37**(6):595–601.
- Hsu CC and Chen CY (2003). Applications of improved grey prediction model for power demand forecasting. *Energy Conversion and Management* **44**(14):2241–2249.
- Hsu CI and Wen YU (1998). Improved grey prediction models for trans-Pacific air passenger market. *Transportation Planning and Technology* **22**(2):87–107.
- Hu YC (2010). Pattern classification by multi-layer perceptron using fuzzy integral-based activation function. *Applied Soft Computing* **10**(3):813–819.
- Hu YC (2016). Pattern classification using grey tolerance rough sets. *Kybernetes* **45**(2):266–281.
- Hu YC, Tzeng GH, Hsu YT and Chen RS (2001). Using learning algorithm to find the developing coefficient and control variable of GM(1,1) model. *Journal of the Chinese Grey System Association* **4**(1):17–26.
- Hu YC, Chiu YJ, Liao YL and Li Q (2015). A fuzzy similarity measure for collaborative filtering using nonadditive grey relational analysis. *Journal of Grey System* **27**(2):93–103.

- Lauret P, Fock E, Randrianarivony RN and Manicom-Ramasamy JF (2008). Bayesian neural network approach to short time load forecasting. *Energy Conversion and Management* **49**(5): 1156–1166.
- Lee SC and Shih LH (2011). Forecasting of electricity costs based on an enhanced gray-based learning model: A case study of renewable energy in Taiwan. *Technological Forecasting and Social Change* **78**(7):1242–1253.
- Lee YS and Tong LI (2011). Forecasting energy consumption using a grey model improved by incorporating genetic programming. *Energy Conversion and Management* **52**(1):147–152.
- Lewis C (1982). *Industrial and Business Forecasting Methods*. London: Butterworth Scientific.
- Liu ZY (2015). *Global Energy Internet*. Beijing: China Electric Power Press.
- Liu S and Lin Y (2006). *Grey Information: Theory and Practical Applications*. London: Springer.
- Luo DR, Guo KZ and Huang HR (2013). Regional economic forecasting combination model based on RAR + SVR. In: Cao BY and Nasser H (Eds) *Fuzzy Information & Engineering and Operations Research & Management, Advances in Intelligent Systems and Computing*, vol. 211. Berlin: Springer, pp. 329–338.
- Makridakis S (1993). Accuracy measures: Theoretical and practical concerns. *International Journal of Forecasting* **9**(4):527–529.
- ManKF, Tang KS and Kwong S (1999). *Genetic Algorithms: Concepts and Designs*. London: Springer.
- Mao MZ and Chirwa EC (2006). Application of grey model GM(1,1) to vehicle fatality risk estimation. *Technological Forecasting and Social Change* **73**(5):588–605.
- National Bureau of Statistics of China (2014). *China Statistical Yearbook 2014*. Beijing: China Statistics Press.
- Pi D, Liu J and Qin X (2010). A grey prediction approach to forecasting energy demand in China. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* **32**(16):1517–1528.
- Rooij AJF, Jain LC and Johnson RP (1996). *Neural Network Training Using Genetic Algorithms*. Singapore: World Scientific.
- Smith KA and Gupta JND (2002). *Neural Networks in Business: Techniques and Applications*. Hershey, PA: Idea Group.
- Suganthi L and Samuel AA (2012). Energy models for demand forecasting—A review. *Renewable and Sustainable Energy Reviews* **16**(2):1223–1240.
- Tsaur RC and Liao YC (2007). Forecasting LCD TV demand using the fuzzy grey model GM(1,1). *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **15**(6):753–767.
- Tutun S, Chou CA and Camyilmaz E (2015). A new forecasting framework for volatile behavior in net electricity consumption: A case study in Turkey. *Energy* **93**(2):2406–2422.
- Vonk E, Jain LC and Johnson RP (1997). *Automatic Generation of Neural Network Architecture Using Evolutionary Computation*. Singapore: World Scientific.
- Wang ZX and Hao P (2016). An improved grey multivariable model for predicting industrial energy consumption in China. *Applied Mathematical Modelling* **40**(11–12):5745–5758.
- Wang CH and Hsu LC (2008). Using genetic algorithms grey theory to forecast high technology industrial output. *Applied Mathematics and Computation* **195**(1):256–263.
- Wei J, Zhou L, Wang F and Wu D (2015). Work safety evaluation in Mainland China using grey theory. *Applied Mathematical Modelling* **39**(2):924–933.
- Weiss SM and Kulikowski CA (1991). *Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems*. CA: Morgan Kaufmann.
- Wen KL (2004). *Grey Systems Modeling and Prediction*. Tucson: Yang's Scientific Research Institute.
- Wu L, Liu S and Wang Y (2012). Grey Lotka–Volterra model and its application. *Technological Forecasting and Social Change* **79**(9):1720–1730.
- Wu L, Liu S, Yao L, Yan S and Liu D (2013). Grey system model with the fractional order accumulation. *Communications in Non-linear Science and Numerical Simulation* **18**(7):1775–1785.
- Wu L, Liu S, Yao L, Yan S and Liu D (2013). The effect of sample size on the grey system model. *Applied Mathematical Modelling* **37**(9):6577–6583.
- Xia C, Wang J and ShortMcMenemy K (2010). Medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks. *Electrical Power and Energy Systems* **32**(7):743–750.
- Xie NM and Liu SF (2009). Discrete grey forecasting model and its optimization. *Applied Mathematical Modelling* **33**(2):1173–1186.
- Zeng B, Meng W and Tong MY (2016). A self-adaptive intelligence grey predictive model with alterable structure and its application. *Engineering Applications of Artificial Intelligence* **50**(C):236–244.
- Zhou P, Ang BW and Poh KL (2006). A trigonometric grey prediction approach to forecasting electricity demand. *Energy* **31**(14):2839–2847.

Received 18 April 2016;  
accepted 13 September 2016