

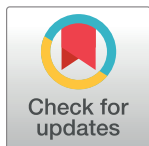
RESEARCH ARTICLE

# A genetic-algorithm-based remnant grey prediction model for energy demand forecasting

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

**1** College of Management & College of Tourism, Fujian Agriculture and Forestry University, Fuzhou City, China, **2** Department of Business Administration, Chung Yuan Christian University, Taoyuan City, Taiwan

\* [ychu@cycu.edu.tw](mailto:ychu@cycu.edu.tw)



## Abstract

Energy demand is an important economic index, and demand forecasting has played a significant role in drawing up energy development plans for cities or countries. As the use of large datasets and statistical assumptions is often impractical to forecast energy demand, the GM(1,1) model is commonly used because of its simplicity and ability to characterize an unknown system by using a limited number of data points to construct a time series model. This paper proposes a genetic-algorithm-based remnant GM(1,1) (GARGM(1,1)) with sign estimation to further improve the forecasting accuracy of the original GM(1,1) model. The distinctive feature of GARGM(1,1) is that it simultaneously optimizes the parameter specifications of the original and its residual models by using the GA. The results of experiments pertaining to a real case of energy demand in China showed that the proposed GARGM(1,1) outperforms other remnant GM(1,1) variants.

## OPEN ACCESS

**Citation:** Hu Y-C (2017) A genetic-algorithm-based remnant grey prediction model for energy demand forecasting. PLoS ONE 12(10): e0185478. <https://doi.org/10.1371/journal.pone.0185478>

**Editor:** Xiaosong Hu, Chongqing University, CHINA

**Received:** March 31, 2017

**Accepted:** September 13, 2017

**Published:** October 5, 2017

**Copyright:** © 2017 Yi-Chung Hu. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper. The dataset made up of historical annual total energy consumption in China were shown in [13].

**Funding:** This research is partially supported by the Ministry of Science and Technology, Taiwan under grant MOST 104-2410-H-033-023-MY2 and MOST 106-2410-H-033-006-MY2. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Energy is necessary for the sustainable development and economic prosperity of a country [1], and this is evidenced by the fact that energy demand has emerged as an important economic index in recent years. With the rapid pace of industrialization, the global demand for energy has increased exponentially in the past decade. Worldwide energy consumption is expected to increase by over 50% before 2030 if the current pattern of global energy consumption continues [2]. Moreover, high energy consumption has a significant and deleterious impact on the environment. This means that the environmental impact of energy consumption will play a crucial role in guiding energy development policies for cities and countries in the future [1]. An important issue in this context is the ability to accurately predict energy demand.

Traditional methods of demand forecasting, including artificial intelligence techniques, multivariate regression, and time series analysis, have been frequently applied to predict energy demand [3–8]. However, a large sample size is usually required to achieve reasonable forecasting accuracy for these methods [9–12]. Furthermore, statistical methods usually require that the data conform to statistical assumptions such as following a particular distribution. However, using large sample sizes or conforming to statistical assumptions is often

impractical [13]. Hence, a forecasting method is needed that can work with small samples without making statistical assumptions to construct an energy demand prediction model [10, 11].

Grey prediction [14] has emerged as a popular technique in the past decade, and is suitable for forecasting energy demand because of its simplicity and ability to characterize an unknown system using a limited number of data points [1]. Grey prediction consists of several forecasting models, of which the GM(1, 1) is commonly used for time series forecasting [15]. The GM(1, 1) model needs only four recent sample data points to achieve reliable and acceptable prediction accuracy [16, 17]. Its effectiveness has been verified through application to a wide range of real-world problems, including energy consumption forecasting [10–13, 18–23], technology management [24, 25], engineering problems [26], optimization model development [27, 28], and general management [29–31].

To increase the forecasting accuracy of the original GM(1, 1) model, further development of the residual GM(1,1) model has been recommended [14, 15]. A residual modification model, also called a remnant GM(1,1) model, is commonly constructed by first building the original GM(1,1) model, and then constructing the residual GM(1,1) model to modify the predicted values obtained by the original model. A number of improved remnant GM(1, 1) models focusing on sign estimation for residual modification have been developed. For instance, Hsu and Chen [32] used a multi-layer perceptron (MLP) to estimate the signs of residual modification to forecast power demand; Hsu [33] used Markov-chain-based sign estimation to modify residuals for the global integrated circuit industry, whereas Lee and Tong [13] combined residual modification with residual genetic programming (GP) sign estimation to develop the GPGM(1, 1) model in view of the importance of forecasting energy demand.

Usually, the original and the residual models are set up separately for remnant models. It would be interesting to investigate whether the prediction accuracy of the traditional remnant GM(1,1) model improves when the GM(1,1) and its residual models are constructed simultaneously. This paper develops a grey forecasting model called the genetic-algorithm-based remnant GM(1,1) model (GARGM(1,1)) with sign estimation that delivers high prediction accuracy. Its distinctive feature is that it can simultaneously optimize the parameters required for the original GM(1,1) and its residual models by a powerful search and optimization method [34–36], the genetic algorithm (GA). This grey prediction model is then applied to forecast energy demand.

The remainder of the paper is organized as follows. Sections 2 and 3 introduce the traditional remnant GM(1,1) and the proposed GARGM(1,1) models, respectively. Section 4 examines the forecasting performance of the GARGM(1,1) model using a dataset collected from China Statistical Yearbook 2008. The results show that the GARGM(1,1) model can outperform other variants of the remnant GM(1,1) model. Section 5 contains a discussion and the conclusions of this study.

## Remnant GM(1,1) model

This section introduces the traditional remnant GM(1,1) model used to improve the predictive accuracy of the original GM(1,1) model. It consists of two main components: the original GM(1,1) model described in Section 2.1, and the residual GM(1,1) model described in Section 2.2.

### Original GM(1,1) model

Let an original data sequence  $\mathbf{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$  be provided by one system and consist of  $n$  samples. A new sequence  $\mathbf{x}^{(1)} = (x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)})$  can then be generated from  $\mathbf{x}^{(0)}$  by the

accumulated generating operation (AGO) [7, 15] as follows:

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

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

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \quad (2)$$

where  $a$  and  $b$  are the developing coefficient and the control variable, respectively. The AGO is used because it can identify potential regularities hidden in the data sequences even if the original data are finite, insufficient, and chaotic.

$\hat{x}_k^{(1)}$ , the predicted value of  $x_k^{(1)}$ , can be obtained by solving the grey difference equation with initial condition  $x_1^{(1)} = x_1^{(0)}$ :

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

$a$  and  $b$  can be estimated by means of a grey difference equation:

$$x_k^{(0)} + az_k^{(1)} = b \quad (4)$$

where the background value  $z_k^{(1)}$  is formulated as follows:

$$z_k^{(1)} = \alpha x_k^{(1)} + (1 - \alpha)x_{k-1}^{(1)} \quad (5)$$

$\alpha$  is usually specified as 0.5 for convenience, but this is not the optimal setting. By using  $n-1$  grey difference equations ( $k = 2, 3, \dots, n$ ),  $a$  and  $b$  can be obtained by the ordinary least-squares method:

$$[a, b]^T = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{y} \quad (6)$$

where

$$\mathbf{B} = \begin{bmatrix} -z_2^{(1)} & 1 \\ -z_3^{(1)} & 1 \\ \vdots & \vdots \\ -z_n^{(1)} & 1 \end{bmatrix} \quad (7)$$

$$\mathbf{y} = [x_2^{(0)}, x_3^{(0)}, \dots, x_n^{(0)}]^T \quad (8)$$

Using the inverse AGO, the predicted value,  $\hat{x}_k^{(0)}$ , of  $x_k^{(0)}$  can be generated as follows:

$$\hat{x}_k^{(0)} = \hat{x}_k^{(1)} - \hat{x}_k^{(1)}, k = 2, 3, \dots, n \quad (9)$$

Therefore,  $\hat{x}_k^{(0)}$  can be formulated as follows:

$$\hat{x}_k^{(0)} = (1 - e^a)(x_1^{(0)} - \frac{b}{a})e^{-a(k-1)}, k = 2, 3, \dots, n \quad (10)$$

Note that  $\hat{x}_1^{(1)} = \hat{x}_1^{(0)}$ .

## Residual GM(1,1) model

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

$$\epsilon_k^{(0)} = |x_k^{(0)} - \hat{x}_k^{(0)}|, k = 2, 3, \dots, n \quad (11)$$

Using the same manner of construction as for the original GM(1,1) model for  $x^{(0)}$ , a residual model can be constructed for  $\epsilon^{(0)}$ . The predicted residual,  $\hat{\epsilon}_k^{(0)}$ , of  $\epsilon_k^{(0)}$  can be derived as follows:

$$\hat{\epsilon}_k^{(0)} = (1 - e^{a_\epsilon})(\epsilon_2^{(0)} - \frac{b_\epsilon}{a_\epsilon})e^{-a_\epsilon(k-1)}, k = 2, 3, \dots, n \quad (12)$$

where  $a_\epsilon$  and  $b_\epsilon$  are the developing coefficient and the control variable respectively. In the remnant GM(1,1) model with sign estimation,  $\hat{x}_k^{(0)}$  is modified by adding  $\hat{\epsilon}_k^{(0)}$  to, or subtracting  $\hat{\epsilon}_k^{(0)}$  from,  $\hat{x}_k^{(0)}$  [37]:

$$\hat{x}_k^{(0)} = (1 - e^a)(x_1^{(0)} - \frac{b}{a})e^{-a(k-1)} + s_k \hat{\epsilon}_k^{(0)}, k = 2, 3, \dots, n \quad (13)$$

where  $s(k)$  denotes the sign (positive or negative) of  $\hat{\epsilon}_k^{(0)}$  with respect to the  $k$ -th year. Compared to the original remnant GM(1,1) model, the sign of each residual in the improved one is unknown and needs to be estimated.

## Genetic-algorithm-based remnant GM(1,1) model

Two main issues need to be addressed in the traditional remnant GM(1,1) model. First, the determination of both the developing coefficient and the control variable, in the original GM(1,1) model and the residual GM(1,1) model, are completely dependent on the background values. However, as the background values cannot be easily determined in advance by decision makers, it is reasonable to try to find developing coefficients and control variables without using background values. The second issues that needs to be addressed is that  $a_\epsilon$  and  $b_\epsilon$  are determined once  $a$  and  $b$  in the original GM(1,1) model have been created. However, in addition to sign estimation, to minimize the difference between the predicted and the actual values, it might be worth examining whether simultaneously determining the four crucial parameters (i.e.,  $a$ ,  $b$ ,  $a_\epsilon$ ,  $b_\epsilon$ ) has an effect on the prediction accuracy of the remnant GM(1,1) model.

The objective of our optimization problem is to minimize the mean absolute percentage error (MAPE) of the training patterns:

$$MAPE = \sum_{k \in TS} \frac{|x_k^{(0)} - \hat{x}_k^{(0)}|}{|TS| \times x_k^{(0)}} \times 100\% \quad (14)$$

where  $TS$  denotes the training or testing data. As the background values are not involved in the formulation of  $\hat{x}_k^{(0)}$ , the computation of MAPE is completely free of the influence of background values. The absolute percentage error (APE), which was used to compare  $\hat{x}_k^{(0)}$  and  $x_k^{(0)}$  with the time series data, was defined as:

$$APE = \frac{|x_k^{(0)} - \hat{x}_k^{(0)}|}{x_k^{(0)}} \times 100\% \quad (15)$$

A method based on the GA is developed to automatically determine the developing coefficients (i.e.,  $a$  and  $a_\epsilon$ ), the control variables (i.e.,  $b$  and  $b_\epsilon$ ), and the sign of the  $k$ -th year (i.e.,  $s_k$ ,  $k = 2, 3, \dots, n$ ) for the improved remnant GM(1,1) model (i.e., GARGM(1,1)). Let  $P_m$  denote the population generated in generation  $m$  ( $1 \leq m \leq n_{max}$ ). Chromosome  $u$  ( $1 \leq u \leq n_{size}$ )

produced in  $P_m$  is represented as  $a_u^m b_u^m a_{\varepsilon u}^m b_{\varepsilon u}^m s_{u,2}^m s_{u,3}^m \dots s_{u,n}^m$  to construct the proposed GARGM(1,1) model, where  $a_u^m$ ,  $b_u^m$ ,  $a_{\varepsilon u}^m$ , and  $b_{\varepsilon u}^m$  are real-valued genes, and  $s_{u,k}^m$  is a binary-valued gene for  $\hat{\varepsilon}_k^{(0)}$ .  $s_{u,k}^m$  has a positive sign when it is one and a negative sign when it is zero.

Let  $n_{size}$  and  $n_{max}$  denote the population size and the maximum number of generations, respectively. Using the MAPE for the training data as the fitness function, having evaluated the fitness value of each chromosome in  $P_m$ , selection, crossover, and mutation are applied until  $n_{size}$  new chromosomes have been generated for  $P_{m+1}$ . The GA can be executed until  $n_{max}$  generations have been generated. The authors of this study performed these genetic operations as described in detail in [38].

## Selection

Using binary tournament selection, two chromosomes from the current population are randomly selected, and the one with the higher fitness is placed in a mating pool. This process is repeated until there are  $n_{size}$  chromosomes in the mating pool.  $n_{size}$  pairs of chromosomes from the pool are then randomly selected for mating. Crossover and mutation operations are applied to a selected parent to reproduce children by altering the chromosomal makeup of the chromosomes of two parents.

## Crossover

For chromosomes  $u (a_u^m b_u^m a_{\varepsilon u}^m b_{\varepsilon u}^m s_{u,2}^m s_{u,3}^m \dots s_{u,n}^m)$  and  $v (a_v^m b_v^m a_{\varepsilon v}^m b_{\varepsilon v}^m s_{v,2}^m s_{v,3}^m \dots s_{v,n}^m)$  ( $1 \leq v \leq n_{size}$ ), each pair of real-valued genes has a crossover probability  $Pr_c$ . The operations are performed as follows:

$$\begin{aligned} a_u^m &= h_1 a_u^m + (1 - h_1) a_v^m, & a_v^m &= h_1 a_v^m + (1 - h_1) a_u^m, \\ b_u^m &= h_2 b_u^m + (1 - h_2) b_v^m, & b_v^m &= h_2 b_v^m + (1 - h_2) b_u^m, \\ a_{\varepsilon u}^m &= h_3 a_{\varepsilon u}^m + (1 - h_3) a_{\varepsilon v}^m, & a_{\varepsilon v}^m &= h_3 a_{\varepsilon v}^m + (1 - h_3) a_{\varepsilon u}^m, \\ b_{\varepsilon u}^m &= h_4 b_{\varepsilon u}^m + (1 - h_4) b_{\varepsilon v}^m, & b_{\varepsilon v}^m &= h_4 b_{\varepsilon v}^m + (1 - h_4) b_{\varepsilon u}^m, \\ s_{u,i}^m &= h_k s_{u,i}^m + (1 - h_k) s_{v,i}^m, & s_{v,i}^m &= h_k s_{v,i}^m + (1 - h_k) s_{u,i}^m, k = 2, \dots, n; \end{aligned}$$

where  $h_1, h_2, h_3, h_4$ , and  $h_k$  are all random numbers in the interval  $[0, 1]$ . In practice, the one-point crossover operation with  $Pr_c$  is used to exchange partial information between binary-valued substrings (i.e.,  $s_{u,2}^m s_{u,3}^m \dots s_{u,n}^m$  and  $s_{v,2}^m s_{v,3}^m \dots s_{v,n}^m$ ) in the selected pair of chromosomes. The crossover point in a substring is chosen randomly. Two new chromosomes are thus generated, and replace their parents in generation  $P_{m+1}$ .

## Mutation

Let  $Pr_m$  denote the probability that mutation is performed for each real-valued parameter in a new chromosome generated by crossover. To avoid excessive perturbation in the gene pool, a low mutation rate should be used. If mutation occurs for a real-valued gene, it is altered by adding a number randomly selected from a specified interval. For each gene of the newly generated binary chromosomes, the mutation operation with  $Pr_m$  is performed on each bit or gene of the string. Each gene in a string can be thus changed either from zero to one or from one to zero with probability  $Pr_m$ .

After crossover and mutation,  $n_{del}$  ( $0 \leq n_{del} \leq n_{size}$ ) chromosomes in  $P_{m+1}$  are randomly removed from the set of new chromosomes (those formed by genetic operations) to make room for additional copies of the chromosome with a maximum fitness value in  $P_m$ . Fig 1 shows a flowchart of the construction of the proposed prediction model using the GA.

## Experimental results

Section 4.1 presents the parameter specifications of the GA-based learning algorithm and Section 4.2 reports the performance of different forecasting methods on a real-world case.

### Parameter specifications of GA

A number of factors can influence the performance of the GA, including population size and the probabilities of applying the crossover and mutation operators. As a matter of fact, no optimal GA parameter specifications exist. The principles recommended by Osyczka [35] and Ishibuchi et al. [36] to specify parameters of GA were as follows:

1. The population size commonly should range from 50 to 500 individuals.
2. The stopping condition should be specified according to the available computation time.
3. Only a small number of elite chromosomes were needed.
4. The crossover probability should be set to a large value because it controls the range of exploration in the solution space.
5. The mutation probability should be set to a small value to avoid generating excessive perturbations.

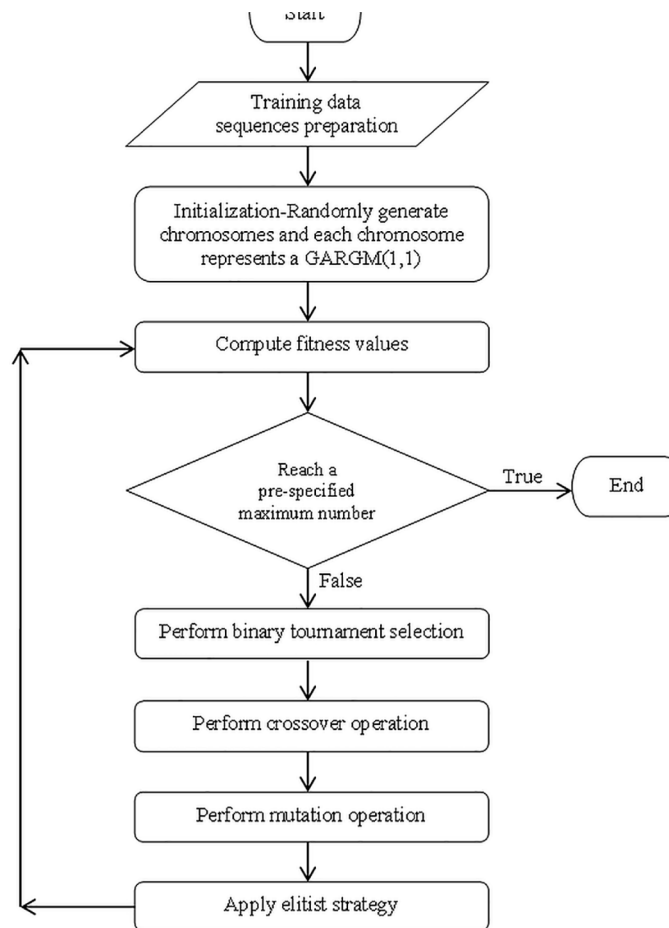
Therefore, the parameters in the experiment were specified as:  $n_{size} = 200$ ,  $n_{max} = 1000$ ,  $n_{del} = 2$ ,  $Pr_c = 0.9$ , and  $Pr_m = 0.01$ .

This experiment constructed the proposed GARGM(1,1) without any complex mechanisms to tune its parameters.

### Application to total energy demand in China

To examine the forecasting capability of the GARGM(1,1) model, an experiment was conducted to compare its performance with the original GM(1,1), the GPGM(1,1), and the improved grey forecasting model using MLP models (MLPGM(1,1)) on a dataset collected from the China Statistical Yearbook 2008. This dataset made up of historical annual total energy consumption in China was shown in [13]. With its rapid economic development and ongoing industrialization, China has played a vital role with regard to energy production and consumption [39]. Indeed, energy demand forecasting has become an increasingly important issue for China [12].

Data from 1990 to 2003 were used for model fitting and those from 2004 to 2007 for ex-post testing. The forecasting results reported in [13] for the original GM(1,1), the MLPGM(1,1), and the GPGM(1,1) models are summarized in Table 1 and illustrated in Fig 2. Table 1 shows that the MAPE values of the original GM(1,1), the MLPGM(1,1), the GPGM(1,1), and the GARGM(1,1) models for the training data were 4.13%, 3.61%, 2.59%, and 1.50%, respectively. Similarly, for the testing data, the MAPE values were 26.21%, 20.23%, 20.23%, and 17.51%, respectively. These results indicate that the GARGM(1,1) model outperformed the other forecasting methods on both training and testing data. The results of ex-post testing were relatively poor for every prediction model because the total energy consumption drastically went up in 2004, as shown in Table 1.



**Fig 1. Flowchart of construction of the proposed prediction model.**

<https://doi.org/10.1371/journal.pone.0185478.g001>

Note that Lewis [40] presented the following MAPE criteria for evaluating a forecasting model:  $MAPE \leq 10$ ,  $10 < MAPE \leq 20$ ,  $20 < MAPE \leq 50$ , and  $MAPE > 50$  correspond to high, good, reasonable and weak forecasting models, respectively. Although the original GM(1,1), the GPGM(1,1), the MLPGM(1,1), and the GARGM(1,1) models all have high forecasting capability on the training data, it is noteworthy that the GARGM(1,1) model exhibited its good forecasting capability, but the original GM(1,1), the GPGM(1,1), and the MLPGM(1,1) models recorded merely reasonable forecasting accuracy values on the testing data. It is hence clear that the GARGM(1, 1) model yielded satisfactory performance compared to the other forecasting methods considered.

## Discussion and conclusions

Energy management is crucial for economic prosperity and environmental security [1]. Energy demand forecasting plays an important role in creating energy policy. GM(1,1) is an appropriate approach to predict energy demand because it uses a limited number of samples to construct a prediction model without statistical assumptions. The present study developed the GARGM(1,1) model to forecast energy demand. The proposed model is sufficiently simple to implement as a computer program. Although the parameter specifications are somewhat subjective, the experimental results showed that they are acceptable.

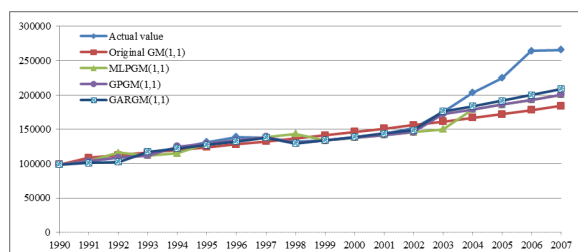


**Table 1. Prediction accuracy obtained by different methods for total energy consumption (unit:  $10^4$  tons of SCE).**

Year	Actual	Original GM(1,1)		MLPGM(1,1)		GPGM(1,1)		GARGM(1,1)	
		Predicted	APE	Predicted	APE	Predicted	APE	Predicted	APE
1990	98703	98703	0	98703	0	98703	0	98703	0
1991	103783	108706.1	4.74	103783	0	103783	0	101195.7	2.49
1992	109170	112335.5	2.9	116225.8	6.46	108445.2	0.66	102596.3	3.14
1993	115993	116086.1	0.08	111804.1	3.61	111804.1	3.61	117195.8	1.04
1994	122737	119962	2.26	115248.8	6.10	124675.1	1.58	121997.9	0.60
1995	131176	123967.2	5.50	129154.8	1.54	129154.8	1.54	127012.8	3.17
1996	138948	128106.2	7.80	133816.1	3.69	133816.1	3.69	132251.6	4.82
1997	137798	132383.3	3.93	138668.2	0.63	138668.2	0.63	137725.9	0.05
1998	132214	136803.3	3.47	143721	8.7	129885.5	1.76	129692.7	1.91
1999	133831	141370.8	5.63	133756.5	0.06	133756.5	0.06	134280.7	0.34
2000	138553	146090.8	5.44	137709.8	0.61	137709.8	0.61	139002.5	0.32
2001	143199	150968.4	5.43	141743.6	1.02	141743.6	1.02	143858.7	0.46
2002	151797	156008.9	2.77	145855.2	3.91	145855.2	3.91	148849.8	1.94
2003	174990	161217.6	7.87	150041.6	14.26	172393.5	1.48	176275.6	0.73
MAPE			4.13		3.61		2.59		1.50
2004	203227	166600.2	18.02	178901.5	11.97	178901.5	11.97	183797.7	9.56
2005	224682	172162.6	23.37	185702.4	17.35	185702.4	17.35	191681.9	14.69
2006	264270	177910.7	32.68	192813.8	27.04	192813.8	27.04	199949.4	24.34
2007	265583	183850.7	30.77	200254.3	24.60	200254.3	24.60	208623.1	21.45
MAPE			26.21		20.23		20.23		17.51

<https://doi.org/10.1371/journal.pone.0185478.t001>

Compared with the MLPGM(1, 1) and the GPGM(1, 1) models, the GARGM(1, 1) model has the advantage of directly determining the developing coefficients and the control variables by the GA without using background values. The required parameters for the RGM are also optimized simultaneously. Experimental results concerning a case of energy demand data from China showed the effectiveness of the proposed forecasting model. In addition to grey prediction models, we examined the prediction performance of two frequently used models, linear regression and the MLP with backpropagation learning. The MLP had an input node, a hidden layer with two neurons, and an output layer with one neuron; it was trained over 10,000 iterations at a learning rate of 0.8. The forecasting results obtained by linear regression and the MLP are summarized in Table 2. It is clear that the prediction accuracy values of linear regression on the training and testing data were 4.20% and 27.76%, respectively [13], whereas those of the MLP on the training and testing data were 3.85% and 18.30%, respectively [20]. Therefore, the proposed GARGM(1,1) outperforms linear regression and the MLP. In case of linear regression, it is reasonable to speculate that the size of the training sample and the



**Fig 2. Predicted values and actual values from 1990 to 2007.**

<https://doi.org/10.1371/journal.pone.0185478.g002>



**Table 2. Prediction accuracy obtained by linear regression and MLP.**

Year	Actual	Linear regression		MLP		GARGM(1,1)	
		Predicted	APE	Predicted	APE	Predicted	APE
1990	98703	101756.6	3.09	93012.6	5.77	98703	0
1991	103783	106243.4	2.37	107674.6	3.75	101195.7	2.49
1992	109170	110730.2	1.43	116921.0	7.10	102596.3	3.14
1993	115993	115217.0	0.67	122130.4	5.29	117195.8	1.04
1994	122737	119703.8	2.47	125034.6	1.87	121997.9	0.60
1995	131176	124190.6	5.33	126861.3	3.29	127012.8	3.17
1996	138948	128677.5	7.39	128373.0	7.61	132251.6	4.82
1997	137798	133164.3	3.36	130080.1	5.60	137725.9	0.05
1998	132214	137651.1	4.11	132407.0	0.15	129692.7	1.91
1999	133831	142137.9	6.21	135788.9	1.46	134280.7	0.34
2000	138553	146624.7	5.83	140696.7	1.55	139002.5	0.32
2001	143199	151111.5	5.53	147565.7	3.05	143858.7	0.46
2002	151797	155598.3	2.5	156595.8	3.16	148849.8	1.94
2003	174990	160085.1	8.52	167469.6	4.30	176275.6	0.73
MAPE			4.20		3.85		1.50
2004	203227	164572.0	19.02	179212.1	11.82	183797.7	9.56
2005	224682	169058.8	24.76	190465.4	15.23	191681.9	14.69
2006	264270	173545.6	34.33	200083.0	24.29	199949.4	24.34
2007	265583	178032.4	32.97	207546.2	21.85	208623.1	21.45
MAPE			27.76		18.30		17.51

<https://doi.org/10.1371/journal.pone.0185478.t002>

statistical assumptions (e.g., homoscedasticity) had a certain impact on the prediction performance of the statistical methods.

Energy demand forecasting can be regarded as a grey system problem [1, 12] because a few factors, such as income and population, influence energy demand. However, how exactly they influence energy demand is unclear. Therefore, based on the superior forecasting performance of the GARGM(1, 1) model in terms of energy demand, the applicability of the proposed forecasting model to other energy forecasting problems, such as electricity consumption in certain developing countries, should be explored. Moreover, sign estimation for the remnant forecasting model should be explored using other artificial intelligence tools such as the functional-link net [41, 42] and other nonadditive neural networks [43–45] could improve prediction accuracy.

Moreover, it has been known that the GA can be time consuming in searching for optimum solutions. Although forecasting energy demand cannot be treated as a kind of large-scale optimization problem, several improved versions, such as the parallel GA [46–50], may be used to expedite the construction of the model or increase its precision for optimum solutions of the proposed GARGM(1,1) model.

## Acknowledgments

The author would like to thank the anonymous referees for their valuable comments. This research is supported by the [Ministry of Science and Technology](#), Taiwan under grant MOST 104-2410-H-033-023-MY2 and MOST 106-2410-H-033-006-MY2. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author Contributions

**Investigation:** Yi-Chung Hu.

**Methodology:** Yi-Chung Hu.

**Validation:** Yi-Chung Hu.

**Writing – original draft:** Yi-Chung Hu.

**Writing – review & editing:** Yi-Chung Hu.

## References

1. Suganthi L, Samuel AA. Energy models for demand forecasting-A review. *Renewable and Sustainable Energy Reviews*. 2012; (16):1223–40.
2. Smith M, Hargroves K, Stasinopoulos P, Stephens R, Desha C, Hargroves S, et al. *Energy Transformed: Sustainable Energy Solutions for Climate Change Mitigation*. Australia: The Natural Edge Project, CSIRO, and Griffith University; 2007.
3. Ediger VS, Akar S. ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy*. 2007; (35):1701–8.
4. Gonzalez PA, Zamarreno JA. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and Buildings*. 2005; (37):595–601.
5. Tutun S, Chou CA. <http://0-www.sciencedirect.com.cylis.lib.cycu.edu.tw/science/article/pii/S0360544215014322> - aff2Canıyılmaz E. A new forecasting framework for volatile behavior in net electricity consumption: A case study in Turkey. *Energy*. 2015; (93):2406–22.
6. Lauret P, Fock E, Randrianarivony RN, Manicom-Ramasamy JF. Bayesian neural network approach to short time load forecasting. *Energy Conversion and Management*. 2008; (49):1156–66.
7. Duran M, Toksari. Estimating the net electricity energy generation and demand using ant colony optimization approach: Case of Turkey. *Energy Policy*. 2009; (37):1181–7.
8. Xia C, Wang J, Short McMenemy K. Medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks. *Electrical Power and Energy Systems*. 2010; (32):743–50.
9. Wang CH, Hsu LC. Using genetic algorithms grey theory to forecast high technology industrial output. *Applied Mathematics and Computation*. 2008; (195):256–63.
10. Feng SJ, Ma YD, Song ZL, Ying J. Forecasting the energy consumption of China by the grey prediction model. *Energy Sources, Part B: Economics, Planning, and Policy*. 2012; (7):376–89.
11. Li DC, Chang CJ, Chen CC, Chen WC. Forecasting short-term electricity consumption using the adaptive grey-based approach-An Asian case. *Omega*. 2012; (40):767–73.
12. Pi D, Liu J, Qin X. A grey prediction approach to forecasting energy demand in China. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*. 2010; (32):1517–28.
13. Lee YS, Tong LI. Forecasting energy consumption using a grey model improved by incorporating genetic programming. *Energy Conversion and Management*. 2011; (52):147–52.
14. Deng JL. Control problems of grey systems. *Systems and Control Letters*. 1982; 1(5):288–94.
15. Liu S, Lin Y. *Grey Information: Theory and Practical Applications*. London: Springer-Verlag; 2006.
16. Wen KL. *Grey Systems Modeling and Prediction*. Tucson: Yang's Scientific Research Institute; 2004.
17. Wang CH, Hsu LC. Using genetic algorithms grey theory to forecast high technology industrial output, *Applied Mathematics and Computation*. 2008; (195):256–63.
18. Hu YC. Electricity consumption forecasting using a neural-network-based grey prediction approach. *Journal of the Operational Research Society*. 2016. <https://doi.org/10.1057/s41274-016-0150-y>
19. Hu YC, Jiang P. Forecasting energy demand using neural-network-based grey residual modification models. *Journal of the Operational Research Society*. 2017; 68(5):556–65.
20. Hu YC. Nonadditive grey prediction using functional-link net for energy demand forecasting. *Sustainability*. 2017; 9(7). <https://doi.org/10.3390/su9071166>
21. Wang ZX, Hao P. An improved grey multivariable model for predicting industrial energy consumption in China. *Applied Mathematical Modelling*. 2016; 40(11–12):5745–58.
22. Xu N, Dang Y, Gong Y. Novel grey prediction model with nonlinear optimized time response method for forecasting of electricity consumption in China. *Energy*. 2017; (118):473–80.

23. Sun W, Xu Y. Research on China's energy supply and demand using an improved Grey-Markov chain model based on wavelet transform. *Energy*. 2017; (118):969–84.
24. Tsaur RC, Liao YC. Forecasting LCD TV demand using the fuzzy grey model GM(1,1). *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 2007; (15):753–67.
25. Ene S, Öztürk N. Grey modelling based forecasting system for return flow of end-of-life vehicles. *Technological Forecasting and Social Change*. 2017; (115):155–66.
26. Chang CJ, Yu L, Jin P. A mega-trend-diffusion grey forecasting model for short-term manufacturing demand. *Journal of the Operational Research Society*. 2016; 67(12):1439–45.
27. Cui J, Liu SF, Zeng BN, Xie NM. A novel grey forecasting model and its optimization. *Applied Mathematical Modelling*. 2013; 37(6):4399–406.
28. Lu JS, Xie WD, Zhou HB, Zhang AJ. An optimized nonlinear grey Bernoulli model and its applications. *Neurocomputing*. 2016; (177):206–14.
29. Mao MZ, Chirwa EC. Application of grey model GM(1,1) to vehicle fatality risk estimation. *Technological Forecasting and Social Change*. 2006; (73):588–605.
30. Wei J, Zhou L, Wang F, Wu D. Work safety evaluation in Mainland China using grey theory. *Applied Mathematical Modelling*. 2015; 39(2):924–33.
31. Hu YC. Predicting foreign tourists for the tourism industry using soft computing-based Grey-Markov models. *Sustainability*. 2017; 9(7), <https://doi.org/10.3390/su9071228>
32. Hsu CC, C. Chen Y. Applications of improved grey prediction model for power demand forecasting. *Energy Conversion and Management*. 2003; (44):2241–9.
33. Hsu LC. Applying the grey prediction model to the global integrated circuit industry. *Technology Forecasting and Social Change*. 2003; 70(6):563–74.
34. Goldberg DE. *Genetic Algorithms in Search, Optimization, and Machine Learning*. MA: Addison-Wesley; 1989.
35. Osyczka A. *Evolutionary Algorithms for Single and Multicriteria Design Optimization*. NY: Physica-Verlag; 2003.
36. Ishibuchi H, Nakashima T, Nii M. *Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining*. Heidelberg: Springer; 2004.
37. Hsu CI, Wen YU. Improved Grey prediction models for trans-Pacific air passenger market. *Transportation Planning and Technology*. 1998; (22):87–107.
38. Hu YC. Recommendation using neighborhood methods with preference-relation-based similarity. *Information Sciences*. 2014; (284):18–30.
39. Liu ZY. *Global Energy Internet*. China Electric Power Press; 2015.
40. Lewis C. *Industrial and Business Forecasting Methods*. London: Butterworth Scientific; 1982.
41. Pao YH. *Adaptive Pattern Recognition and Neural Networks*. Reading: Addison-Wesley; 1989.
42. Pao YH. Functional-link net computing: Theory, system architecture, and functionalities. *Computer*. 1992; 25(5):76–9.
43. Hu YC, Tseng FM. Functional-link net with fuzzy integral for bankruptcy prediction. *Neurocomputing*. 2007; 70(16–18):2959–68.
44. Hu YC, Wang JH, Chang CY. Flow-based grey single-layer perceptron with fuzzy integral. *Neurocomputing*. 2012; (91):86–9.
45. Chiang JH. Choquet fuzzy integral-based hierarchical networks for decision analysis. *IEEE Transactions on Fuzzy Systems*. 1999; 7(1):63–71.
46. Lim DD, Ong YS, Jin YC, Sendho B, Lee BS. Efficient hierarchical parallel genetic algorithms using grid computing. *Future Generation Computer Systems*. 2007; 23(4):658–70.
47. Cantu-Paz E, Goldberg DE. *Efficient parallel genetic algorithms: theory and practice*. *Computer Methods in Applied Mechanics and Engineering*. 2000; 186(2–4):221–38.
48. Tsoulos IG, Tzallas A, Tsalikakis D. PDoublePop: An implementation of parallel genetic algorithm for function optimization. *Computer Physics Communications*. 2016; (209):183–9.
49. Yu WJ, Li JZ, Chen WN, Zhang J. A parallel double-level multiobjective evolutionary algorithm for robust optimization. *Applied Soft Computing*. 2017; (59):258–75.
50. Liu YY, Wang S. A scalable parallel genetic algorithm for the Generalized Assignment Problem. *Parallel Computing*. 2015; (46):98–119.