

Article

Nonadditive Grey Prediction Using Functional-Link Net for Energy Demand Forecasting

Yi-Chung Hu ^{1,2}

¹ College of Management & College of Tourism, Fujian Agriculture and Forestry University, Fuzhou 350002, China; ychu@cycu.edu.tw

² Department of Business Administration, Chung Yuan Christian University, Chung Li Dist., Taoyuan 32023, Taiwan

Received: 2 June 2017; Accepted: 27 June 2017; Published: 3 July 2017

Abstract: Energy demand prediction plays an important role in sustainable development. The GM(1,1) model has drawn our attention to energy demand forecasting because it only needs a few data points to construct a time series model without statistical assumptions. Residual modification is often considered as well to improve the accuracy of predictions. Several residual modification models have been proposed, but they focused on residual sign estimation, whereas the FLNGM(1,1) model using functional-link net (FLN) can estimate the sign as well as the modification range for each predicted residual. However, in the original FLN, an activation function with an inner product assumes that criteria are independent of each other, so additivity might influence the forecasting performance of FLNGM(1,1). Therefore, in this study, we employ the FLN with a fuzzy integral instead of an inner product to propose a nonadditive FLNGM(1,1). Experimental results based on real energy demand cases demonstrate that the proposed grey prediction model performs well compared with other grey residual modification models that use sign estimation and the additive FLNGM(1,1).

Keywords: energy demand; grey prediction; neural network; fuzzy integral; residual modification

1. Introduction

Several advanced countries have worked on the green new deal to promote the development of green energy technology, and to achieve the ultimate goal of sustainable development. Sustainable development is based on three main traits, including energy security, economic development, and environmental protection [1]. These traits involve the amount of energy demand and supplies, effective use of the energy for industries, and development of new technologies for renewable energy. Environmental impacts due to energy consumption are inevitable because economic development is ongoing, and will have a significant role when devising the policy of sustainable development for cities or countries [2]. In the last few decades, continuous economic development and population increases worldwide have led to rapid growth in the demand for energy. If the current global energy consumption pattern continues, the global energy consumption will increase by over 50% before 2030 [3]. In terms of the global electricity demand, the average annual growth rate was 2.6% from 1990 to 2000 and 3.3% from 2000 to 2010. It is expected to reach 2.8% from 2010 to 2020 [4]. Since the amount of energy demand must be taken into account first for sustainable development, this leads to an important issue related to how to accurately predict energy demand.

Many forecasting methods, including artificial intelligence techniques, multivariate regression, and time series analysis, have frequently been applied to energy demand forecasting [5–13]. A large number of samples are required for multivariate regression and time series analysis like the autoregressive integrated moving average (ARIMA) [14]. The performance of the above-mentioned methods can be significantly affected by the number and the representativeness of observations [15,16]. However, using long-term data to build energy consumption prediction models may be impractical

because the average annual growth rate of energy consumption is high and unstable. Beyond this, statistical methods usually require that the data conform to statistical assumptions, such as having a normal distribution [17], yet energy consumption data often do not conform to these usual statistical assumptions [17], limiting the forecasting capabilities of statistical methods. Therefore, to construct an energy demand prediction model, a forecasting method is needed that works well with small samples and without making any statistical assumptions [18]. One of the grey prediction models, GM(1,1), has drawn our attention to energy demand forecasting [2]. Indeed, how factors such as income and population influence the demand for energy is not clear, so energy demand forecasting can be regarded as a grey system problem [2,16].

The GM(1,1) model only needs four recent sample data points to achieve a reliable and acceptable accuracy of prediction [14]. Several versions have been proposed to improve the accuracy of its predictions [19–23]. In addition, a residual modification model built on residuals obtained from the original GM(1,1) may be an effective solution [24,25]. In terms of residual modification, in addition to residual sign estimation [17,26–28], Hu [29] proposed a new model called FLNGM(1,1) which uses the functional-link net (FLN) with an effective function approximation capability [30–33] to estimate both the sign and the modification range of the predicted residuals obtained from the residual GM(1,1) model.

In the original FLNGM(1,1), the hyperbolic tangent function is considered as the activation function. Clearly, the output of this function uses a weighted sum of the connection weights with an enhanced pattern. It is assumed that the additivity property of the interaction among individual features results in an enhanced pattern. However, the attributes are not always independent of each other [34–37], so an assumption of additivity for the enhanced pattern may not be reasonable [38], thereby affecting the forecasting performance of FLNGM(1,1). The Choquet fuzzy integral [39–41] does not assume the independence of each criterion, so in this study we propose a nonadditive residual modification model, called nonadditive FLNGM(1,1) (N-FLNGM(1,1)) for energy demand forecasting, where we replace the weighted sum with the Choquet integral inside the hyperbolic tangent function. A genetic algorithm (GA) [42] is used to construct the proposed nonadditive FLNGM(1,1) model with high prediction accuracy.

The remainder of this paper is organized as follows. Section 2 introduces the traditional GM(1,1) model using residual modification with sign estimation. Sections 3 and 4 present residual modification using FLN and the proposed N-FLNGM(1,1) model, respectively. Section 5 examines the energy demand forecasting performance of the N-FLNGM(1,1) model based on real energy demand cases in China. In Section 6, we discuss the outcomes and give our conclusions.

2. GM(1,1) Model Using Residual Modification with Sign Estimation

2.1. Original GM(1,1) Model

Given an original data sequence $\mathbf{X}^{(0)} = (x_1^{(0)}, \dots, x_n^{(0)})$ made up of n samples, then a new sequence, $\mathbf{X}^{(1)} = (x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)})$, can be generated from $x_1^{(0)}$ by the accumulated generating operation [9,25] as follows:

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

and $x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}$ can then be 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 $\hat{x}_k^{(1)}$ can be obtained by solving the differential equation with an initial condition that $x_1^{(1)} = x_1^{(0)}$:

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

and thus $\hat{x}_k^{(1)} = x_1^{(0)}$ holds. Then, a and b can be estimated by a grey difference equation:

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

where $z_k^{(1)}$ is the background value,

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

and α is usually specified as 0.5 for convenience. In turn, a and b can be obtained using the ordinary least-squares method:

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

where

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

and

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

Using the inverse accumulated generating operation, the predicted value of $x_k^{(0)}$ is

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

Therefore,

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

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

2.2. Residual Modification with Sign Estimation

To build a residual modification model, the original GM(1,1) is constructed first, followed by the residual GM(1,1). Let $\varepsilon^{(0)} = (\varepsilon_2^{(0)}, \varepsilon_3^{(0)}, \dots, \varepsilon_n^{(0)})$ denote the sequence of absolute residual values, where

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

Using the same construction for $\mathbf{X}^{(\circ)}$ as the original GM(1,1) model, a residual model can be established for $\varepsilon^{(0)}$, and the predicted residual of $\varepsilon_k^{(0)}$ is

$$\hat{\varepsilon}_k^{(0)} = (1 - e^{a_\varepsilon}) (\varepsilon_2^{(0)} - \frac{b_\varepsilon}{a_\varepsilon}) e^{-a_\varepsilon(k-1)}, k = 3, 4, \dots, n \quad (12)$$

In the remnant GM(1,1) model with sign estimation, $\hat{x}_k^{(0)}$ is modified to $\hat{x}_{k^r}^{(0)}$ by adding or subtracting $\hat{\varepsilon}_k^{(0)}$ from $\hat{x}_k^{(0)}$ [27].

$$\hat{x}_{k^r}^{(0)} = \hat{x}_k^{(0)} + s_k \hat{\varepsilon}_k^{(0)}, k = 2, 3, \dots, n \quad (13)$$

where s_k denotes the positive or negative sign of $\hat{\varepsilon}_k^{(0)}$ with respect to the k -th year. Those residual sign estimation methods mentioned above can be used to determine s_k .

3. Residual Modification Using FLN

A single-layer feed-forward FLN is an appropriate tool for obtaining sign and range estimations due to its effective function approximation ability. An enhanced pattern with respect to a single input denoted by $t \in \Re$ can be generated as $(t, \sin(\pi t), \cos(\pi t), \sin(2\pi t), \cos(2\pi t))$ through a functional link [31,32], where t denotes a specific time point, such as the t -th year.

In the original FLNGM(1,1), the hyperbolic tangent function is used as the activation function in the output node,

$$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (14)$$

where $\tanh(z)$ lies within the range $(-1, 1)$. Let θ be the bias to the output node. When $(t, \sin(\pi t), \cos(\pi t), \sin(2\pi t), \cos(2\pi t))$ is presented to the net, z can be computed as follows:

$$z = w_1 t + w_2 \sin(\pi t) + w_3 \cos(\pi t) + w_4 \sin(2\pi t) + w_5 \cos(2\pi t) + \theta \quad (15)$$

The actual output value, y , is just to equal $\tanh(z)$, and y can be interpreted as the range of modification for $\hat{x}_k^{(0)}$. Let y_k denote the actual output value with respect to $\hat{x}_k^{(0)}$. Then, $y_k = 1$ indicates that $\hat{x}_k^{(0)}$ can be modified as its tolerable upper limit, whereas $y_k = -1$ denotes that $\hat{x}_k^{(0)}$ can be modified as its tolerable lower limit. An advantage compared with residual estimation methods is that the modification range of the original $\hat{x}_k^{(0)}$ is not restricted to $\hat{\varepsilon}_k^{(0)}$. Thus, it can be seen that the additivity assumption of the interaction among $t, \sin(\pi t), \cos(\pi t), \sin(2\pi t)$, and $\cos(2\pi t)$ holds; i.e., these terms are assumed to be independent of each other.

In particular, based on the idea of three-sigma limits, which are used to set the upper and lower control limits in statistical quality control charts [43], the value $\hat{x}_{k\text{FLN}}^{(0)}$ predicted by the proposed model is formulated heuristically as follows.

$$\hat{x}_{k\text{FLN}}^{(0)} = (1 - e^a) \left(x_1^{(0)} - \frac{b}{a} \right) e^{-a(k-1)} + 3y_k \hat{\varepsilon}_k^{(0)}, \quad k = 2, 3, \dots, n \quad (16)$$

where $3\hat{\varepsilon}_k^{(0)}$ refers to data within three residuals with respect to $\hat{x}_k^{(0)}$ and represents the tolerable maximum range for modifying $\hat{x}_k^{(0)}$.

4. Nonadditive Residual Modification Model

Let $(t, \sin(\pi t), \cos(\pi t), \sin(2\pi t), \cos(2\pi t))$ be represented by $(v_1, v_2, v_3, v_4, v_5)$, and let $X = \{v_1, v_2, v_3, v_4, v_5\}$, where X is called the feature space. A fuzzy measure is a nonadditive set function that can be used along with a fuzzy integral for aggregating information sources [36]. Among various fuzzy measures, the λ -fuzzy measure has been suggested for computation of the fuzzy integral because of its convenience users [44–46]. The advantage is that, after determining fuzzy densities $\mu_1, \mu_2, \dots, \mu_5$, where μ_k denotes $\mu(\{v_k\})$, λ can be uniquely determined from the condition $\mu(X) = 1$. $\mu(E_k)$ can be computed as follows:

$$\mu(E_k) = \frac{1}{\lambda} \left[\prod_{i=kn} (1 + \lambda \mu_i) - 1 \right] \quad (17)$$

where $E_k = \{v_k, v_{k+1}, \dots, v_5\}$ ($1 \leq k \leq 5$). As illustrated in Figure 1, fuzzy densities can be used as connection weights, and they can be determined automatically by the GA.

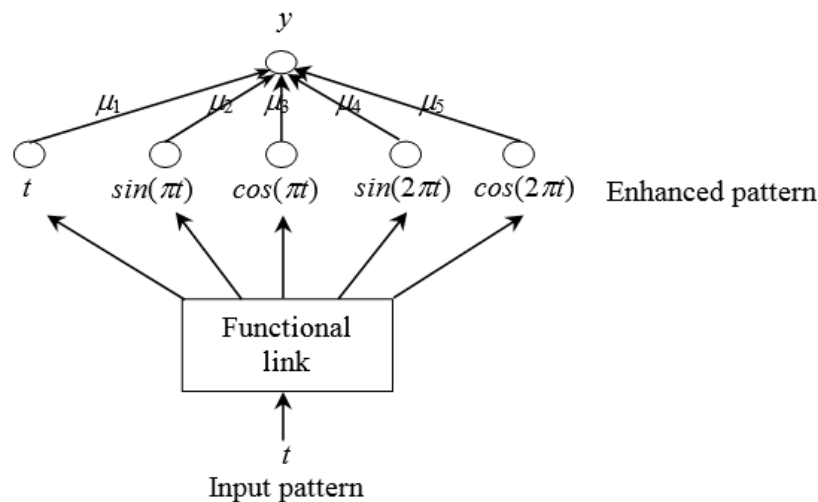


Figure 1. A functional-link net with Choquet fuzzy integral.

Let f be a non-negative, real-valued measurable function defined on X . The element in X with $\min\{f(v_k) | k = 1, 2, \dots, 5\}$ is renumbered as one, where $f(v_k)$ denotes the performance value of v_k . In other words, all the elements v_k are rearranged in order of descending $f(v_k)$, such that $f(v_1) \geq f(v_2) \geq \dots \geq f(v_5)$. The Choquet integral $(c) \int f d\mu$ over X of f with respect to μ is defined as follows.

$$(c) \int f d\mu = \sum_{k=1}^5 f(v_k) [\mu(E_k) - \mu(E_{k+1})] \quad (18)$$

where $\mu(E_6)$ is specified as zero. The Choquet integral illustrated in Figure 2 comprises five different rectangular areas. Since the assumption of additivity among features is not warranted in real applications, it is reasonable to use the Choquet integral with respect to aggregate $t, \sin(\pi t), \cos(\pi t), \sin(2\pi t)$, and $\cos(2\pi t)$ to form the nonadditive hyperbolic tangent function denoted by $(c)\tanh(z)$. This definitely differentiates N-FLNGM(1,1) from the original FLNGM(1,1) models.

In the construction of the original FLNGM(1,1), a real-valued GA is designed to automatically determine the connection weights (i.e., $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$) and the bias (i.e., θ) for the N-FLNGM(1,1). Indeed, Rooij et al. [47] noted that GAs are appropriate tools for training perceptron-like networks. To construct the model with high prediction accuracy, the mean absolute percentage error (MAPE) for the training patterns is considered. MAPE can be treated as the benchmark, and it is more stable than the more commonly used mean absolute error and root mean square error [48,49]. MAPE is formulated as follows:

$$\text{MAPE} = \sum_{k \in TS} \frac{|x_k^{(0)} - \hat{x}_{k^{\text{FLN}}}^{(0)}|}{|TS| \times x_k^{(0)}} \times 100\% \quad (19)$$

where TS denotes the training or testing data. Ref. [50] presented 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. It is reasonable to define the fitness function as MAPE. Thus, the fitness of each individual is specified by a single objective in the combinatorial optimization problem.

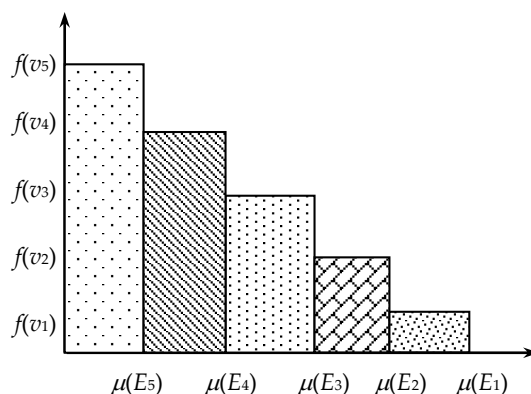


Figure 2. Graphical representation of Choquet fuzzy integral.

Let n_{size} and n_{max} denote the population size and maximum number of generations, respectively, and P_m denote the population generated in generation m ($1 \leq m \leq n_{max}$). After evaluating the fitness value for 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 performed until n_{max} generations have been generated. When the stopping condition has been satisfied, the algorithm is terminated, and the best chromosome with maximum fitness value among all successive generations serves as the desired solution to examine the generalization ability of the proposed N-FLNGM(1,1) model. The selection, crossover, and mutation operations required for N-FLNGM(1,1) are similar to those for FLNGM(1,1), so the details of the GA are omitted for simplicity and they can be found in [29].

5. Experimental Results

In Asia, China has become increasingly influential in terms of energy production and consumption [4]. It is interesting to conduct experiments on real energy demand cases from China to compare the energy demand forecasting ability of the proposed N-FLNGM(1,1) model with that of the additive FLNGM(1,1) and other grey residual modification models that use sign estimation, including the residual modification model using multi-layer perceptron (MLPGM(1,1)) [26], genetic programming (GPGM(1,1)) [17], and Markov chain [27]. The prediction accuracy of different prediction models were demonstrated by two real cases related to the energy demand in China. To ensure a fair comparison with the original FLNGM(1,1), the parameter specifications used for GA were the same as those described by Hu (2016), including $n_{size} = 200$, $n_{max} = 1000$, probabilities for crossover and mutation are 0.7 and 0.01 respectively.

5.1. Case I

An experiment was conducted based on the historical annual energy demand data from China collected between 1990 and 2007. As reported by Lee and Tong [17], the data from 1990 to 2003 were used for model fitting, and those from 2004 to 2007 were used for ex-post testing. The forecasting results obtained by Lee and Tong [17] using the original GM(1,1), MLPGM(1,1), and GPGM(1,1) models are summarized in Table 1 and illustrated in Figure 2.

Table 1 and Figure 3 show the prediction performance of the different forecasting models. The MAPE for model fitting obtained by N-FLNGM(1,1) was slightly inferior to that by the original FLNGM(1,1). However, we can see that N-FLNGM(1,1) outperformed the other forecasting models compared with the testing data. Using the testing data, in terms of MAPE, both the original FLNGM(1,1) and the proposed N-FLNGM(1,1) had good forecasting abilities, whereas the original GM(1,1), GPGM(1,1), MLPGM(1,1), and Markov-chain sign estimation only had reasonably fair forecasting abilities. A massive change occurred up to 2004, which may explain why the ex-post testing results were not as good as the model-fitting results.

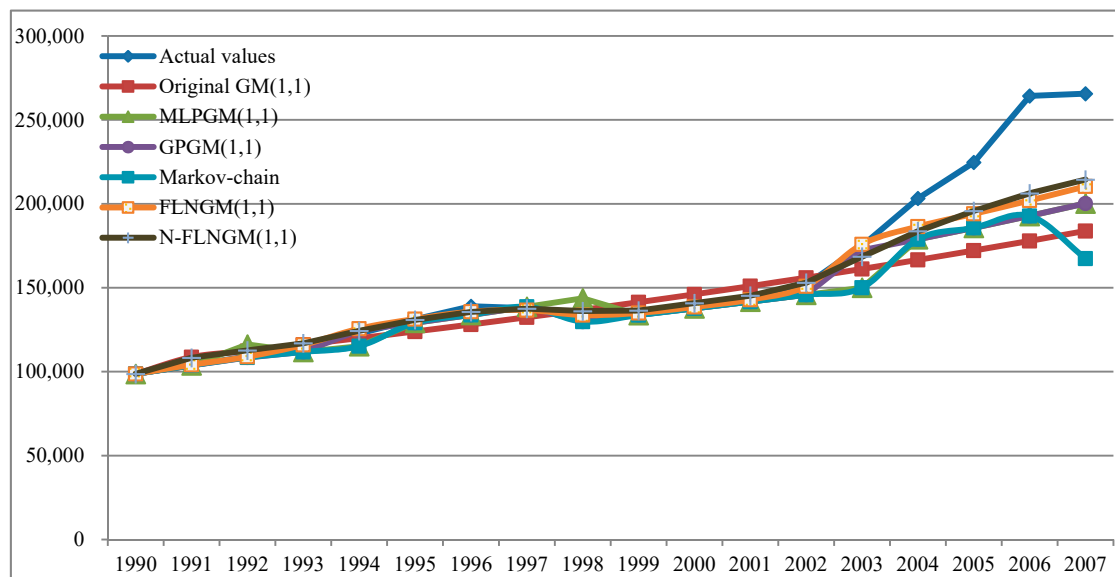


Figure 3. Predicted and actual values of different forecasting models for Case I.

5.2. Case II

The second experiment was conducted on the historical annual electricity demand of China, collected from China Statistical Yearbook 2014 [51]. As described by Zhou et al. [52], data from 1981 to 1998 were used for model fitting, and the other data for testing. The forecasting results obtained by the different forecasting models are shown in Table 2 and Figure 4. We can see that the MAPE for the original GM(1,1), MLPGM(1,1), GPGM(1,1), Markov-chain sign estimation, original FLNGM(1,1), and the proposed N-FLNGM(1,1) using the training data were 2.28%, 2.03%, 1.44%, 2.31%, 0.10%, and 0.13%, respectively. Using the testing data, the MAPE were 7.24%, 3.90%, 3.90%, 3.90%, 1.52%, and 1.41%, respectively. Figure 5 illustrates the APE of the different forecasting models. Based on these results, it is obvious that the proposed N-FLNGM(1,1) model yielded comparable performance to the other forecasting models considered. Additionally, Zhou et al. [52] showed that the MAPE for an autoregressive integrated moving average model and trigonometric grey prediction model were 3.25% and 2.12%, respectively, for training data, and 2.64% and 2.37% for testing data, which are inferior to the results obtained by the proposed N-FLNGM(1,1) model.

Table 1. Prediction accuracy obtained by different forecasting models for energy demand (unit: 10^4 tons of standard coal equivalent (SCE)). MAPE: mean absolute percentage error.

Year	Actual	GM(1,1)		MLPGM(1,1)		GPGM(1,1)		Markov-Chain		FLNGM(1,1)		N-FLNGM(1,1)	
		Predicted	APE	Predicted	APE	Predicted	APE	Predicted	APE	Predicted	APE	Predicted	APE
1990	98,703	98,703	0	98,703	0	98,703	0	98,703	0	98,703	0	98,703	0
1991	103,783	108,706.1	4.74	103,783	0	103,783	0	103,783	0	104,116.2	0.32	108,332.5	4.38
1992	109,170	112,335.5	2.9	116,225.8	6.46	108,445.2	0.66	108,445.2	0.66	108,913.6	0.23	112,708.3	3.24
1993	115,993	116,086.1	0.08	111,804.1	3.61	111,804.1	3.61	111,804.1	3.61	115,973.2	0.02	116,938.2	0.82
1994	122,737	119,962	2.26	115,248.8	6.10	124,675.1	1.58	115,248.8	6.10	125,629.8	2.36	124,309.9	1.28
1995	131,176	123,967.2	5.50	129,154.8	1.54	129,154.8	1.54	129,154.8	1.54	131,380.7	0.16	130,842	0.25
1996	138,948	128,106.2	7.80	133,816.1	3.69	133,816.1	3.69	133,816.1	3.69	135,703.7	2.33	135,615.4	2.40
1997	137,798	132,383.3	3.93	138,668.2	0.63	138,668.2	0.63	138,668.2	0.63	136,710.1	0.79	137,554.2	0.18
1998	132,214	136,803.3	3.47	143,721	8.7	129,885.5	1.76	129,885.5	1.76	133,735.8	1.15	136,059.7	2.92
1999	133,831	141,370.8	5.63	133,756.5	0.06	133,756.5	0.06	133,756.5	0.06	135,005.6	0.88	136,387.9	1.92
2000	138,553	146,090.8	5.44	137,709.8	0.61	137,709.8	0.61	137,709.8	0.61	138,603.2	0.04	140,826.2	1.65
2001	143,199	150,968.4	5.43	141,743.6	1.02	141,743.6	1.02	141,743.6	1.02	142,946.8	0.18	145,273.3	1.45
2002	151,797	156,008.9	2.77	145,855.2	3.91	145,855.2	3.91	145,855.2	3.91	150,954.8	0.55	153,016.6	0.81
2003	174,990	161,217.6	7.87	150,041.6	14.26	172,393.5	1.48	150,041.6	14.26	175,848.7	0.49	168,529.3	3.69
MAPE			4.13		3.61		2.59		2.70		0.68		1.78
2004	203,227	166,600.2	18.02	178,901.5	11.97	178,901.5	11.97	178,901.5	11.97	186,450.1	8.26	183,696.6	9.61
2005	224,682	172,162.6	23.37	185,702.4	17.35	185,702.4	17.35	185,702.4	17.35	194,058.0	13.63	195,771.8	12.87
2006	264,270	177,910.7	32.68	192,813.8	27.04	192,813.8	27.04	192,813.8	27.04	202,011.1	23.56	206,254.6	21.95
2007	265,583	183,850.7	30.77	200,254.3	24.60	200,254.3	24.60	167,447.1	36.95	210,377.6	20.79	214,426.9	19.26
MAPE			26.21		20.23		20.23		23.22		16.56		15.92

Table 2. Prediction accuracy obtained by different forecasting models for electricity demand (unit: 100 million kWh).

Year	Actual	GM(1,1)		MLPGM(1,1)		GPGM(1,1)		Markov-chain		FLNGM(1,1)		N-FLNGM(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	3280	0	3214.3	2.00	3374.4	2.88
1983	3519	3611.5	2.63	3638.0	3.38	3585.0	1.88	3585.0	1.88	3548.5	0.84	3625.0	3.01
1984	3778	3919.5	3.75	3888.3	2.92	3888.3	2.92	3888.3	2.92	3845.3	1.78	3893.6	3.06
1985	4118	4253.9	3.30	4217.1	2.41	4217.1	2.41	4217.1	2.41	4166.4	1.18	4189.4	1.73
1986	4507	4616.7	2.43	4573.3	1.47	4573.3	1.47	4573.3	1.47	4513.6	0.15	4523.1	0.36
1987	4985	5010.5	0.51	4959.4	0.51	4959.4	0.51	4959.4	0.51	4889.0	1.93	4897.4	1.76
1988	5467	5437.9	0.53	5377.6	1.63	5498.2	0.57	5377.6	1.63	5294.7	3.15	5320.3	2.68
1989	5865	5901.7	0.63	5830.7	0.59	5830.7	0.59	5830.7	0.59	5733.0	2.25	5775.0	1.53
1990	6230	6405.1	2.81	6321.4	1.47	6321.4	1.47	6321.4	1.47	6209.5	0.33	6276.4	0.74
1991	6775	6951.4	2.60	6852.8	1.15	6852.8	1.15	6852.8	1.15	6777.9	0.04	6859.8	1.25
1992	7542	7544.3	0.03	7428.1	1.51	7428.1	1.51	7428.1	1.51	7596.7	0.72	7545.1	0.04
1993	8426.5	8187.8	2.83	8050.8	4.46	8324.8	1.21	8050.8	4.46	8425.2	0.02	8319.0	1.28
1994	9260.4	8886.2	4.04	8724.8	5.78	9047.6	2.30	8724.8	5.78	9207.4	0.57	9144.2	1.25
1995	10,023.4	9644.1	3.78	9834.3	1.89	9834.3	1.89	9453.9	5.68	10,000.4	0.23	10,000.1	0.23
1996	10,764.3	10,466.7	2.76	10,690.9	0.68	10,690.9	0.68	10,242.5	4.85	10,742.7	0.20	10,685.5	0.73
1997	11,284.4	11,359.5	0.67	11,623.7	3.01	11,095.3	1.68	11,095.3	1.68	11,273.8	0.09	11,264.8	0.17
1998	11,598.4	12,328.4	6.29	12,017.1	3.61	12,017.1	3.61	12,017.1	3.61	11,796.7	1.71	11,866.7	2.31
MAPE			2.28		2.03		1.44		2.31		0.10		0.13
1999	12,305.2	13,379.9	8.73	13,013	5.75	13,013.0	5.75	13,013.0	5.75	12,581.9	2.25	12,627.3	2.62
2000	13,471.4	14,521.2	7.79	14,088.8	4.58	14,088.8	4.58	14,088.8	4.58	13,535.1	0.47	13,565.9	0.70
2001	14,633.5	15,759.8	7.70	15,250.3	4.21	15,250.3	4.21	15,250.3	4.21	14,598.4	0.24	14,710.7	0.53
2002	16,331.5	17,104	4.73	16,503.6	1.05	16,503.5	1.05	16,503.5	1.05	15,821.4	3.12	16,040.2	1.78
MAPE			7.24		3.90		3.90		3.90		1.52		1.41

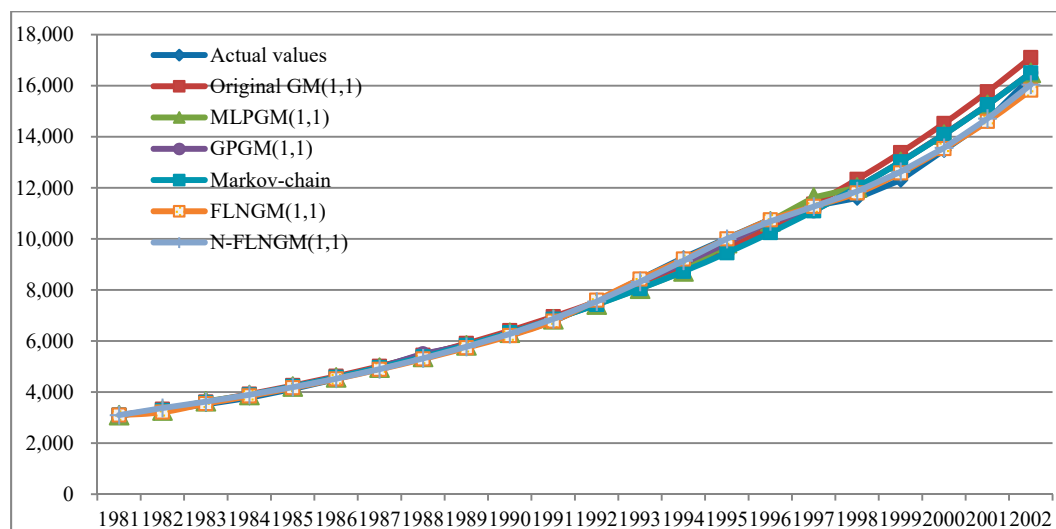


Figure 4. Predicted and actual values of different forecasting models for Case II.

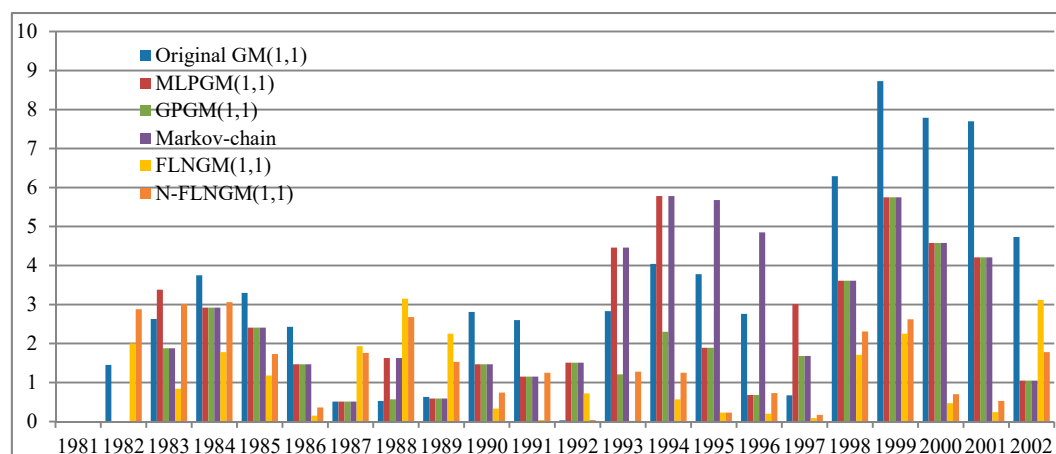


Figure 5. APE of different forecasting models for Case II.

6. Discussion and Conclusions

The GM(1,1) model is the most frequently used grey prediction model, and it has played an important role in energy demand forecasting because it only requires a limited number of samples to construct a prediction model without statistical assumptions. In this paper, the proposed N-FLNGM(1,1) is built on the original FLNGM(1,1) model, which uses FLN to estimate the sign and modification range for each predicted residual simultaneously by using the predicted value obtained from the GM(1,1) model. We performed experiments to validate the effectiveness of the proposed N-FLNGM(1,1) model for energy demand forecasting. Experimental results demonstrate that the proposed nonadditive grey prediction model performs well compared with the additive FLNGM(1,1). In addition to grey prediction models, we further examine the prediction performance of a frequently-used prediction model—multi-layer perceptron (MLP) with backpropagation learning. MLP here has one input node, one hidden layer with two neurons, and one output layer with one neuron, training with 10,000 iterations, and learning rate of 0.8. The forecasting results obtained by MLP for case I and II are summarized in Tables 3 and 4, respectively. It can be seen that the proposed N-FLNGM(1,1) outperforms MLP.

Table 3. Prediction accuracy obtained by multi-layer perceptron (MLP) and autoregressive integrated moving average (ARIMA) for case I.

Year	Actual	MLP		N-FLNGM(1,1)	
		Predicted	APE	Predicted	APE
1990	98,703	93,012.6	5.77	987,03	0
1991	103,783	107,674.6	3.75	108,332.5	4.38
1992	109,170	116,921.0	7.10	112,708.3	3.24
1993	115,993	122,130.4	5.29	116,938.2	0.82
1994	122,737	125,034.6	1.87	124,309.9	1.28
1995	131,176	126,861.3	3.29	130,842	0.25
1996	138,948	128,373.0	7.61	135,615.4	2.40
1997	137,798	130,080.1	5.60	137,554.2	0.18
1998	132,214	132,407.0	0.15	136,059.7	2.92
1999	133,831	135,788.9	1.46	136,387.9	1.92
2000	138,553	140,696.7	1.55	140,826.2	1.65
2001	143,199	147,565.7	3.05	145,273.3	1.45
2002	151,797	156,595.8	3.16	153,016.6	0.81
2003	174,990	167,469.6	4.30	168,529.3	3.69
MAPE			3.85		1.78
2004	203,227	179,212.1	11.82	183,696.6	9.61
2005	224,682	190,465.4	15.23	195,771.8	12.87
2006	264,270	200,083.0	24.29	206,254.6	21.95
2007	265,583	207,546.2	21.85	214,426.9	19.26
MAPE			18.30		15.92

Table 4. Prediction accuracy obtained by MLP and ARIMA for case II.

Year	Actual	MLP		N-FLNGM(1,1)	
		Predicted	APE	Predicted	APE
1981	3096	2947.5	4.80	3096	0
1982	3280	3255.4	0.75	3374.4	2.88
1983	3519	3573.7	1.55	3625.0	3.01
1984	3778	3900.4	3.24	3893.6	3.06
1985	4118	4234.9	2.84	4189.4	1.73
1986	4507	4579.0	1.60	4523.1	0.36
1987	4985	4938.1	0.94	4897.4	1.76
1988	5467	5323.1	2.63	5320.3	2.68
1989	5865	5752.7	1.91	5775.0	1.53
1990	6230	6253.5	0.38	6276.4	0.74
1991	6775	6855.1	1.18	6859.8	1.25
1992	7542	7575.5	0.44	7545.1	0.04
1993	8426.5	8396.8	0.35	8319.0	1.28
1994	9260.4	9252.6	0.08	9144.2	1.25
1995	10,023.4	10,051.4	0.28	10,000.1	0.23
1996	10,764.3	10,724.4	0.37	10,685.5	0.73
1997	11,284.4	11,250.6	0.30	11,264.8	0.17
1998	11,598.4	11,646.4	0.41	11,866.7	2.31
MAPE			1.34		0.13
1999	12,305.2	11,942.1	2.95	12,627.3	2.62
2000	13,471.4	12,166.3	9.69	13,565.9	0.70
2001	14,633.5	12,341.1	15.67	14,710.7	0.53
2002	16,331.5	12,481.5	23.57	16,040.2	1.78
MAPE			12.97		1.41

The proposed N-FLNGM(1,1) model is totally different from the original FLNGM(1,1), MLPGM(1,1), GPGM(1,1), and Markov-chain sign estimation. First, when the Choquet fuzzy integral with respect to a λ -fuzzy measure is incorporated into the FLN to consider the interaction among

features in the enhanced pattern, the testing results obtained by the additive FLNGM(1,1) can be further improved by the proposed N-FLNGM(1,1). Second, MLPGM(1,1), GPGM(1,1), and Markov-chain sign estimation contributed to estimate residual sign s_k in Equation (13) to improve the prediction accuracy of the original GM(1,1) model. The common characteristic of these three models is that $\hat{x}_k^{(0)}$ can be adjusted by either $\hat{\varepsilon}_k^{(0)}$ or $-\hat{\varepsilon}_k^{(0)}$. This is too tight a restriction for the modification range. In contrast, originating from the idea of three-sigma limits, N-FLNGM(1,1) presents a novel updated rule for $\hat{x}_k^{(0)}$ by FLN, where the adjustment of $\hat{x}_k^{(0)}$ ranges between $-3\hat{\varepsilon}_k^{(0)}$ and $3\hat{\varepsilon}_k^{(0)}$. It can be seen that N-FLNGM(1,1) outperformed MLPGM(1,1), GPGM(1,1), and Markov-chain sign estimation compared with data used for model fitting and ex-post testing. It should be noted that in a perceptron-like net, it is not easy to explain the meaning of the connection weights [53]. However, an advantage of the proposed N-FLNGM(1,1) compared with the original FLNGM(1,1) is that μ_k can be interpreted as the degree of importance of v_k .

Over the next two decades, coal, natural gas, and crude oil can be the main energy supplies driving the world economy. It is expected that the growth of the crude oil demand will mainly come from emerging markets from 2015 to 2035, and one half of the growth will be in China [54]. As a matter of fact, energy consumption in China has been mainly provided by coal and crude oil. For instance, the China Statistical Yearbook 2014 [51] showed that approximately two-thirds of the total energy consumed was provided by coal and 18% by oil in 2013. This leads to inevitable environmental impacts on China. Undoubtedly, energy demand prediction has become increasingly important when devising sustainable development plans for China [16]. The proposed N-FLNGM(1,1) has demonstrated its potential for energy demand forecasting.

When it comes to fuzzy integrals, the Sugeno integral is the other most commonly used fuzzy integral. However, only the maximum and minimum operators are involved in the Sugeno integral, so the Choquet integral is preferable to the Sugeno integral for many decision problems [36], which is why we use the Choquet integral rather than the Sugeno integral. In order to perform a soft aggregation, two special ordered weighted averaging (OWA) operators—S-OWA-OR and S-OWA-AND—can be employed to replace the maximum and minimum operators, respectively [55]. In future research, it would be interesting to examine the forecasting ability of a nonadditive prediction model using the Sugeno integral combined with OWA operators.

In this study, both the original and the residual GM(1,1) models use the least squares method to obtain the developing coefficient and control variable, which depend on the background value. However, it is not easy to determine the background value. Hu et al. [56] presented a novel neural-network-based GM(1,1) model (NNGM(1,1)) to resolve this troublesome problem by automatically determining the developing coefficient and control variable. Thus, it would be interesting to examine whether incorporating NNGM(1,1) into N-FLNGM(1,1) instead of the traditional GM(1,1) model might affect the prediction performance of N-FLNGM(1,1) in energy demand forecasting.

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

Conflicts of Interest: The author declare no conflict of interest.

References

1. Taiwan Bureau of Energy. White Paper on Energy and Industrial Technology, Technical Report. Available online: https://www.moeaboe.gov.tw/ecw/populace/images/file_icon/pdf.png (accessed on 30 May 2017).
2. Suganthi, L.; Samuel, A.A. Energy models for demand forecasting—A review. *Renew. Sustain. Energy Rev.* **2012**, *16*, 1223–1240. [CrossRef]
3. Smith, M.; Hargroves, K.; Stasinopoulos, P.; Stephens, R.; Desha, C.; Hargroves, S. *Energy Transformed: Sustainable Energy Solutions for Climate Change Mitigation*; The Natural Edge Project, CSIRO, and Griffith University: Brisbane, Australia, 2007.
4. Liu, Z.Y. *Global Energy Internet*; China Electric Power Press: Beijing, China, 2015.

5. Boroojeni, K.G.; Amini, M.H.; Bahrami, S.; Iyengar, S.S.; Sarwat, A.I.; Karabasoglu, O. A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon. *Electr. Power Syst. Res.* **2017**, *142*, 58–73. [[CrossRef](#)]
6. Ediger, V.S.; Akar, S. ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy* **2007**, *35*, 1701–1708. [[CrossRef](#)]
7. Lauret, P.; Fock, E.; Randrianarivony, R.N.; Manicom-Ramasamy, J.F. Bayesian neural network approach to short time load forecasting. *Energy Convers. Manag.* **2000**, *49*, 1156–1166. [[CrossRef](#)]
8. Mohanty, S.; Patra, P.K.; Sahoo, S.S.; Mohanty, A. Forecasting of solar energy with application for a growing economy like India: Survey and implication. *Renew. Sustain. Energy Rev.* **2017**, *78*, 539–553. [[CrossRef](#)]
9. Toksari, M.D. Estimating the net electricity energy generation and demand using ant colony optimization approach: Case of Turkey. *Energy Policy* **2009**, *37*, 1181–1187. [[CrossRef](#)]
10. Tutun, S.; Chou, C.A.; Canıylmaz, E. A new forecasting framework for volatile behavior in net electricity consumption: A case study in Turkey. *Energy* **2015**, *93*, 2406–2422. [[CrossRef](#)]
11. Verdejo, H.; Awerkin, A.; Becker, C.; Olguin, G. Statistic linear parametric techniques for residential electric energy demand forecasting. A review and an implementation to Chile. *Renew. Sustain. Energy Rev.* **2017**, *74*, 512–521. [[CrossRef](#)]
12. Xia, C.; Wang, J.; McMenemy, K.S. Medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks. *Electr. Power Energy Syst.* **2010**, *32*, 743–750. [[CrossRef](#)]
13. Yang, Y.; Chen, Y.; Wang, Y.; Li, C.; Li, L. Modelling a combined method based on ANFIS and neural network improved by DE algorithm: A case study for short-term electricity demand forecasting. *Appl. Soft Comput.* **2016**, *49*, 663–675. [[CrossRef](#)]
14. Wang, C.H.; Hsu, L.C. Using genetic algorithms grey theory to forecast high technology industrial output. *Appl. Math. Comput.* **2008**, *195*, 256–263. [[CrossRef](#)]
15. Li, D.C.; Chang, C.J.; Chen, C.C.; Chen, W.C. Forecasting short-term electricity consumption using the adaptive grey-based approach—An Asian case. *Omega* **2012**, *40*, 767–773. [[CrossRef](#)]
16. Pi, D.; Liu, J.; Qin, X. A grey prediction approach to forecasting energy demand in China, Energy Sources, Part A: Recovery. *Util. Environ. Eff.* **2010**, *32*, 1517–1528.
17. Lee, Y.S.; Tong, L.I. Forecasting energy consumption using a grey model improved by incorporating genetic programming. *Energy Convers. Manag.* **2011**, *52*, 147–152. [[CrossRef](#)]
18. Feng, S.J.; Ma, Y.D.; Song, Z.L.; Ying, J. Forecasting the energy consumption of China by the grey prediction model, Energy Sources, Part B: Economics. *Plan. Policy* **2012**, *7*, 376–389.
19. Chen, Y.; He, K.; Zhang, C. A novel grey wave forecasting method for predicting metal prices. *Resour. Policy* **2016**, *49*, 323–331. [[CrossRef](#)]
20. Li, K.; Liu, L.; Zhai, J.; Khoshgoftaar, T.M.; Li, T. The improved grey model based on particle swarm optimization algorithm for time series prediction. *Eng. Appl. Artif. Intell.* **2016**, *55*, 285–291. [[CrossRef](#)]
21. Wang, Z.X.; Hao, P. An improved grey multivariable model for predicting industrial energy consumption in China. *Appl. Math. Model.* **2016**, *40*, 5745–5758. [[CrossRef](#)]
22. Zeng, B.; Meng, W.; Tong, M.Y. A self-adaptive intelligence grey predictive model with alterable structure and its application. *Eng. Appl. Artif. Intell.* **2016**, *50*, 236–244. [[CrossRef](#)]
23. Zhao, H.; Guo, S. An optimized grey model for annual power load forecasting. *Energy* **2016**, *107*, 272–286. [[CrossRef](#)]
24. Deng, J.L. Control problems of grey systems. *Syst. Control Lett.* **1982**, *1*, 288–294.
25. Liu, S.; Lin, Y. *Grey Information: Theory and Practical Applications*; Springer-Verlag: London, UK, 2006.
26. Hsu, C.C.; Chen, C.Y. Applications of improved grey prediction model for power demand forecasting. *Energy Convers. Manag.* **2003**, *44*, 2241–2249. [[CrossRef](#)]
27. Hsu, C.I.; Wen, Y.U. Improved Grey prediction models for trans-Pacific air passenger market. *Transp. Plan. Technol.* **1998**, *22*, 87–107. [[CrossRef](#)]
28. Hsu, L.C. Applying the grey prediction model to the global integrated circuit industry. *Technol. Forecast. Soc. Chang.* **2003**, *70*, 563–574. [[CrossRef](#)]
29. Hu, Y.C. Grey prediction with residual modification using functional-link net and its application to energy demand forecasting. *Kybernetes* **2017**, *46*, 349–363. [[CrossRef](#)]
30. Hu, Y.C. Functional-link nets with genetic-algorithm-based learning for robust nonlinear interval regression analysis. *Neurocomputing* **2009**, *72*, 1808–1816. [[CrossRef](#)]

31. Pao, Y.H. *Adaptive Pattern Recognition and Neural Networks*; Addison-Wesley: Reading, Boston, MA, USA, 1989.
32. Pao, Y.H. Functional-link net computing: Theory, system architecture, and functionalities. *Computer* **1992**, *25*, 76–79. [CrossRef]
33. Park, G.H.; Pao, Y.H. Unconstrained word-based approach for off-line script recognition using density-based random-vector functional-link net. *Neurocomputing* **2000**, *31*, 45–65. [CrossRef]
34. Hu, Y.C. Nonadditive similarity-based single-layer perceptron for multi-criteria collaborative filtering. *Neurocomputing* **2014**, *129*, 306–314. [CrossRef]
35. Hu, Y.C.; Chiu, Y.J.; Liao, Y.L.; Li, Q. A fuzzy similarity measure for collaborative filtering using nonadditive grey relational analysis. *J. Grey Syst.* **2015**, *27*, 93–103.
36. Wang, Z.; Leung, K.S.; Klir, G.J. Applying fuzzy measures and nonlinear integrals in data mining. *Fuzzy Sets Syst.* **2005**, *156*, 371–380. [CrossRef]
37. Wang, W.; Wang, Z.; Klir, G.J. Genetic algorithms for determining fuzzy measures from data. *J. Intell. Fuzzy Syst.* **1998**, *6*, 171–183.
38. Hu, Y.C.; Tseng, F.M. Functional-link net with fuzzy integral for bankruptcy prediction. *Neurocomputing* **2007**, *70*, 2959–2968. [CrossRef]
39. Murofushi, T.; Sugeno, M. An interpretation of fuzzy measure and the Choquet integral as an integral with respect to a fuzzy measure. *Fuzzy Sets Syst.* **1989**, *29*, 201–227. [CrossRef]
40. Murofushi, T.; Sugeno, M. A theory of fuzzy measures: Representations, the Choquet integral, and null sets. *J. Math. Anal. Appl.* **1991**, *159*, 532–549. [CrossRef]
41. Murofushi, T.; Sugeno, M. Some quantities represented by the Choquet integral. *Fuzzy Sets Syst.* **1993**, *56*, 229–235. [CrossRef]
42. Goldberg, D.E. *Genetic Algorithms in Search, Optimization, and Machine Learning*; Addison-Wesley: Massachusetts, MA, USA, 1989.
43. Montgomery, D.C. *Statistical Quality Control*; John Wiley & Sons: New Jersey, NJ, USA, 2005.
44. Kuncheva, L.I. *Fuzzy Classifier. Design*; Physica-Verlag: Heidelberg, Germany, 2000.
45. Sugeno, M. Fuzzy measures and fuzzy integrals: A survey. In *Fuzzy Automata and Decision Processes*; Gupta, M.M., Saridis, G.N., Gaines, B.R., Eds.; North Holland: New York, NY, USA, 1977; pp. 89–102.
46. Sugeno, M.; Narukawa, Y.; Murofushi, Y. Choquet integral and fuzzy measures on locally compact space. *Fuzzy Sets Syst.* **1998**, *7*, 205–211. [CrossRef]
47. Rooij, A.J.F.; Jain, L.C.; Johnson, R.P. *Neural Network Training Using Genetic Algorithms*; World Scientific: Hackensack, NJ, USA, 1996.
48. Lee, S.C.; Shih, L.H. Forecasting of electricity costs based on an enhanced gray-based learning model: A case study of renewable energy in Taiwan. *Technol. Forecast. Soc. Chang.* **2011**, *78*, 1242–1253. [CrossRef]
49. Makridakis, S. Accuracy measures: Theoretical and practical concerns. *Int. J. Forecast.* **1993**, *9*, 527–529. [CrossRef]
50. Lewis, C. *Industrial and Business Forecasting Methods*; Butterworth Scientific: London, UK, 1982.
51. National Bureau of Statistics of China. *China Statistical Yearbook 2014*; China Statistics Press: Beijing, China, 2014.
52. Zhou, P.; Ang, B.W.; Poh, K.L. A trigonometric grey prediction approach to forecasting electricity demand. *Energy* **2006**, *31*, 2839–2847. [CrossRef]
53. Eric, R.; Seema, S. Bankruptcy prediction by neural network. In *Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real-World Performance*; Trippi, R.R., Turban, E., Eds.; McGraw-Hill: Chicago, IL, USA, 1996; pp. 243–259.
54. British Petroleum. Energy Outlook, Technical Report. Available online: <https://www.bp.com/content/dam/bp/pdf/energy-economics/energy-outlook-2017/bp-energy-outlook-2017.pdf> (accessed on 30 May 2017).
55. Yager, R.R. Elements selection from a fuzzy subset using the fuzzy integral, IEEE Transactions on Systems, Man Cybern. **1993**, *23*, 467–477. [CrossRef]
56. Hu, Y.C.; Tzeng, G.H.; Hsu, Y.T.; Chen, R.S. Using learning algorithm to find the developing coefficient and control variable of GM(1,1) model. *J. Chin. Grey Syst. Assoc.* **2001**, *4*, 17–26.

