sustainability

MDPI

Article

Predicting Foreign Tourists for the Tourism Industry Using Soft Computing-Based Grey–Markov Models

Yi-Chung Hu [1,2]

[1] College of Management & College of Tourism, Fujian Agriculture and Forestry University, Fuzhou City 35000, China; ychu@cycu.edu.tw; Tel.: +886-3-2655130

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

Received: 7 June 2017; Accepted: 7 July 2017; Published: 13 July 2017

Abstract: Accurate prediction of foreign tourist numbers is crucial for each country to devise sustainable tourism development policies. Tourism time series data often have significant temporal fluctuation, so Grey–Markov models based on a grey model with a first order differential equation and one variable, GM(1,1), can be appropriate. To further improve prediction accuracy from Grey–Markov models, this study incorporates soft computing techniques to estimate a modifiable range for a predicted value, and determine individual state bounds for the Markov chain. A new residual value is formulated by summing the transition probability matrices with different steps. The proposed grey prediction model was applied to foreign tourist forecasting using historical annual data collected from Taiwan Tourism Bureau and China National Tourism Administration. The experimental results show that the proposed grey prediction model performs well in comparison with other Grey–Markov models considered.

Keywords: foreign tourist; soft computing; grey prediction; Markov chain; neural network

---

1. Introduction

Development of the global tourism industry has contributed significantly to economic flourishing for a country. In 2016, the World Travel and Tourism Council estimated a 3.1% growth rate for the tourism industry, which was larger than the estimated global GDP growth (2.3%). The global tourism industry contributes significantly to employment, providing 107.83 million jobs or 3.6% of total employment in 2015, and will account for 135.88 million jobs by 2026. By 2026, capital investment is estimated to be USD 1254.2 billion, with international tourist arrivals expected to be 1.93 billion, generating expenditure of USD 2056.0 billion. Thus, the foreign visitors expenditure contributes much to the global tourism industry.

Accurate prediction of foreign tourist numbers has become crucial for governments to be able to set up relevant sustainable tourism development and marketing strategies to promote the tourism industry. National authorities should carefully consider the changing number of foreign tourists. The variety of international tourism has raised a challenging task for foreign tourist prediction [1]. Grey prediction models [2] have drawn much attention because they can characterize an unknown system from limited data [3–5], without requiring conformance to statistical assumptions, such as normal distributions. The widely used grey model with a first order differential equation and one variable, GM(1,1), for example, can be set up using only four recent sample data points [6–13].

A residual model is often constructed to improve GM(1,1) prediction [3,14] where the predicted residuals can be used to adjust GM(1,1) predicted values. Many studies have demonstrated the Grey–Markov model, denoted by MCGM(1,1), can significantly improve prediction accuracy over the original GM(1,1) [1,15–20]. MCGM(1,1) uses GM(1,1) to identify the trend of historical data, and then applies the Markov chain to correct the residuals. Hsu and Wen [21] and Hsu [22] used

Markov chain sign estimation to modify residuals for the air passenger market and global integrated circuit industry, respectively. Hsu et al. [16] proposed a Markov Fourier model to forecast stock market turning time. Kumar and Jain [18] applied MCGM(1,1) to predict conventional energy consumption. Li et al. [19] combined the regression model with Markov chain for thermal electric power generation. Mao and Sun [23] applied MCGM(1,1) for fire accident prediction. Sun et al. [1] proposed a MCGM(1,1) variant using the Cuckoo search algorithm for foreign tourist arrival prediction, and Wang [24] showed MCGM(1,1) effectiveness for tourism demand prediction. Xie et al. [20] proposed a novel Markov model to estimate the probability that one energy component could transit to another energy component.

The combination of grey prediction and soft computing can better represent system dynamics with uncertainty and nonlinearity [16]. Thus, this paper concentrates on building an effective MCGM(1,1) model multi-step transitions for predicting foreign tourist numbers, incorporating soft computing. The proposed model explicitly considered the following issues. First, the upper and lower bounds of individual states for the Markov chain are usually required to be known in advance. However, these troublesome bounds are not easy to determine beforehand. Therefore, we adopted the genetic algorithm (GA), a powerful optimization method [25–28], to maximize prediction accuracy for the MCGM(1,1) model.

Second, the original MCGM(1,1) usually uses the whole of a Markov chain predicted residual to modify a predicted value from GM(1,1). However, this restriction could impact on MCGM(1,1) prediction accuracy. Therefore, we require a modifiable range rather than using the whole range. Since this involves a connection between time periods and modifiable ranges, we employed the functional link net (FLN) with effective function approximation [29–32] to estimate a modifiable range for a time period. Finally, the developing coefficient and control variable associated with the original GM(1,1) are usually determined by the background value. However, this background values is difficult to determine accurately. On the other hand, neural network based GM(1,1), denoted by NNGM(1,1), [14,33] using a single-layer perceptron (SLP) can avoid the requirement for the background value. Therefore, the current study incorporated NNGM(1,1) into MCGM(1,1). Therefore, this study proposes a novel soft computing based MCGM(1,1) (SC-MCGM(1,1)).

The remainder of the paper is organized as follows. Section 2 introduces the NNGM(1,1) model and Section 3 presents the proposed SC-MCGM(1,1) model. Section 4 describes the construct of the proposed model using GA, and Section 5 examines the model performance using two real cases of foreign tourist forecasting. Section 6 provides our conclusions. This paper is concluded with Section 6.

2. NNGM(1,1) for Generating Predicted Values

Using the accumulated generating operation (AGO) [3], a new sequence $\mathbf{x}^{(1)} = (x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)})$ can be generated from an original data sequence $\mathbf{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)})$,

$$x_k^{(1)} = \sum_{j=1}^{k} x_k^{(0)}, k = 1, 2, \ldots n \tag{1}$$

and $x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)}$ are approximated by the first-order differential equation,

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

where a is the developing coefficient and b is the control variable. AGO also helps identify regularity hidden in data sequences, even if the collected data are finite, insufficient, and chaotic.

The predicted value, $\hat{x}_k^{(1)}$, associated with $x_k^{(1)}$ can be derived from the differential equation with the initial condition $x_1^{(1)} = x_1^{(0)}$,

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

so $\hat{x}_1^{(1)} = x_1^{(0)}$ holds. The predicted value of $x_k^{(0)}$ can be obtained by using the inverse accumulated generating operation,

$$\hat{x}_k^{(0)} = \hat{x}_k^{(1)} - \hat{x}_{k-1}^{(1)}, k = 2, 3, \ldots n. \tag{4}$$

Therefore,

$$\hat{x}_k^{(0)} = (1 - e^a)(x_1^{(0)} - \frac{b}{a})e^{-a(k-1)}, k = 2, 3, \ldots, n \tag{5}$$

where a and b can be estimated from the grey difference equation

$$x_k^{(0)} + a\, z_k^{(1)} = b \tag{6}$$

where $z_k^{(1)}$ is the background value. However, $z_k^{(1)}$ is not easily determined. Therefore, to obtain a and b without requiring $z_k^{(1)}$, an NNGM(1,1) model was established using a single layer perceptron (SLP) accompanied by the cost function

$$E(a, b) = \frac{1}{2}\sum_k (x_k^{(0)} - \hat{x}_k^{(0)})^2, k = 2, 3, \ldots, n \tag{7}$$

where a and b serve as connection weights for the SLP, and the learning rules can be easily derived by using the gradient descent method on $E(a,b)$. For further NNGM(1,1) details, the reader is referred to [33].

3. The Proposed SC-MCGM(1,1) Model

3.1. Generating Transition Probability Matrices

For the proposed residual modification model, we applied the Markov chain to modify the residuals produced by the NNGM(1,1). Let $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ denote the sequence of residual values obtained from training data, where

$$\varepsilon_k = \left| x_k^{(0)} - \hat{x}_k^{(0)} \right|, k = 1, 2, \ldots, n \tag{8}$$

$[\varepsilon_{\min}, \varepsilon_{\max}]$ denotes the residual range, where $\varepsilon_{\min}$ and $\varepsilon_{\max}$ are the minimum and maximum values of ε_k, respectively, and $[\varepsilon_{\min}, \varepsilon_{\max}]$ can be divided into r intervals ($r \geq 2$) with each interval treated as a state. The actual state of ε_k, denoted by s_k, can be determined depending on where it locates. $r - 1$ partition points, $p_1, p_2, \ldots, p_{r-1}$, can be defined for r intervals, where $\varepsilon_{\min} < p_1 < p_2 < \ldots < p_{r-1} < \varepsilon_{\max}$.

Subsequently, an m-step transition probability matrix $P^{(m)}$ can be generated from training patterns as

$$P^{(m)} = \begin{bmatrix} p_{11}^{(m)} & p_{12}^{(m)} & \cdots & p_{1r}^{(m)} \\ p_{21}^{(m)} & p_{22}^{(m)} & \cdots & p_{2r}^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r1}^{(m)} & p_{r2}^{(m)} & \cdots & p_{rr}^{(m)} \end{bmatrix} \tag{9}$$

where $p_{ij}^{(m)}$ represents the transition probability of going from state i to state j ($1 \leq i, j \leq r$) by m steps,

$$p_{ij}^{(m)} = \frac{t_{ij}^{(m)}}{t_i} \tag{10}$$

where $t_{ij}^{(m)}$ represents the number of transitions of going from state i to state j by m steps; and t_i represents the amount of state i among the sequence of relative errors. $p_{ii}^{(m)}$ can be specified directly as 1 when the sum of elements in the row i equals zero. In other words, such a state is treated as an absorbing state.

3.2. Determining Centers for Individual States

Sun et al. [1] recommended that the number of intervals can be defined by Sturge formula [34]:

$$r = \frac{\ln(n)}{\ln 2} \tag{11}$$

Let c_w ($1 \leq w \leq r$) be the representative point of state w, whose lower and upper bounds are l_w and u_w, respectively. For convenience, c_w is traditionally formulated as

$$c_w = \frac{l_w + u_w}{2} \tag{12}$$

Nevertheless, it was more reasonable to formulate c_w as [1,19]

$$c_w = \alpha_w\, l_w + (1 - \alpha_w)\, u_w \tag{13}$$

where $0 \leq \alpha_w \leq 1$.

3.3. Computing Predicted Residual Values

Let $s_k^{(m)}$ denote the state that corresponds to m transitions ahead of s_k associated with ε_k. To determine the predicted state $\hat{s}_k$ for time period k, at most k-1 ($k \geq 2$) transitions ahead can be considered. The actual states including $s_k^{(1)}$, $s_k^{(2)}$, ... , and $s_k^{(m)}$ can be incorporated into the determination of $\hat{s}_k$ if $k \geq m + 1$. In contrast, only $s_k^{(1)}$, $s_k^{(2)}$, ... , and $s_k^{(k-1)}$ are considered if $k < m + 1$. For instance, to determine $\hat{s}_3$, one and two transition steps from s_2 and s_1, respectively, can be used for two transitions ahead of $\hat{s}_3$ (i.e., m = 2), and s_2 and s_1 are $s_3^{(1)}$ and $s_3^{(2)}$, respectively. Then, $\hat{s}_3$ can be simply determined from s_2 and s_1 even though $m > 2$.

Let $p_{s_k^{(m)}} = (p_{s_k^{(m)},1},\ p_{s_k^{(m)},2}, \ldots, p_{s_k^{(m)},r})$ denote the row vector in $P^{(m)}$ corresponding to $s_k^{(m)}$. If $k \geq m + 1$, then we sum $p_{s_k^{(1)}}$, $p_{s_k^{(2)}}$, ... , and $p_{s_k^{(m)}}$ for m previous transitions,

$$v_{kl} = p_{s_k^{(1)},l} + p_{s_k^{(2)},l} + \ldots + p_{s_k^{(m)},l} \tag{14}$$

where $1 \leq l \leq r$. Otherwise,

$$v_{kl} = p_{s_k^{(1)},l} + p_{s_k^{(2)},l} + \ldots + p_{s_k^{(k-1)},l} \tag{15}$$

To explain v_{kl} as the degree in [0,1] to which $\hat{\varepsilon}_k$ locates in state l,

$$v_{kl} = \frac{v_{kl}}{m} \tag{16}$$

Traditionally, state l can be directly assigned to $\hat{s}_k$ associated with the predicted residual value $\hat{\varepsilon}_k$ [29,34] when

$$v_{kl} = \max_{i=1..r} v_{ki} \tag{17}$$

Then, state l is the reachable state with the maximum likelihood for $\hat{\varepsilon}_k$. In such a case, $\hat{\varepsilon}_k$ just equals c_l. However, in addition to c_l, c_i ($i \neq l$) can also contribute to $\hat{\varepsilon}_k$ if $v_{ki} \neq 0$. Therefore, considering the contribution from different representative points for $\hat{\varepsilon}_k$, v_{kl} can be normalized as

$$v_{kl} = \frac{v_{kl}}{\sum\limits_{i=1}^{r} v_{ki}} \tag{18}$$

and $\hat{\varepsilon}_k$ can becomes

$$\hat{\varepsilon}_k = v_{k1}c_1 + v_{k2}c_2 + \ldots + v_{kr}c_r \tag{19}$$

3.4. FLN for Determining New Predicted Values

We apply $\hat{\varepsilon}_k$ obtained from the Markov chain to adjust $\hat{x}_k^{(0)}$, calculating the new predicted value as

$$\widetilde{x}_k^{(0)} = \hat{x}_k^{(0)} + y_k\hat{\varepsilon}_k^{(m)},\ k = 1, 2, \ldots, n \tag{20}$$

where y_k ranges from -1 to 1 and can be interpreted as the degree to which $\hat{x}_k^{(0)}$ can be adjusted. That is, if y_k is positive, then larger y_k means it is more likely that $\hat{x}_k^{(0)}$ will be adjusted toward $\hat{x}_k^{(0)} + \hat{\varepsilon}_k^{(m)}$, whereas, if y_k is negative, smaller y_k means it is more likely $\hat{x}_k^{(0)}$ will be adjusted toward $\hat{x}_k^{(0)} - \hat{\varepsilon}_k^{(m)}$.

How to obtain y_k is left to FLN. Let $t_k \in \Re$ denote the time period k with respect to $\hat{x}_k^{(0)}$. For one variable x, a FLN with functional-expansion expansion like {x, sin(πx), cos(πx), sin($2\pi x$), cos($2\pi x$),...} is effective to approximate a nonlinear function associated with x [29,30,35]. In principle, the components in the functional expansion representation can be unrestrictedly extended for x, but this is not practical. However, (t_k, sin(πx), cos(πx), sin($2\pi x$), cos($2\pi x$), sin($4\pi x$))) with respect to x is acceptable [30,31]. Using this pattern, the corresponding actual output value can be obtained from the output node as

$$y_k = \tanh(w_1t_k + w_2\sin(\pi t_k) + w_3\cos(\pi t_k) + w_4\sin(2\pi t_k) + w_5\cos(2\pi t_k) + w_6\sin(4\pi t_k) + \theta) \tag{21}$$

where w_1, w_2, w_3, w_4, w_5, and w_6 are connection weights; tanh denotes the hyperbolic tangent function; and θ is the bias to the output node.

4. A Genetic Algorithm for Constructing the SC-MCGM(1,1)

The problem of constructing the SC-MCGM(1,1) with high prediction accuracy can be formulated as maximizing the reciprocal of MAPE for training data,

$$\text{MAPE} = \sum_{k \in TS} \frac{\left| x_k^{(0)} - \widetilde{x}_k^{(0)} \right|}{|TS| \times x_k^{(0)}} \times 100\% \tag{22}$$

where TS denotes training or testing data. To minimize MAPE, a real-valued GA was developed to automatically determine 6 + 2r parameters that are not easily directly accessed, including the connection weights (w_1, w_2, w_3, w_4, w_5, w_6), bias (θ), partition points (p_1, p_2, ... , p_{r-1}), and relative weights in respective intervals (α_1, α_2, ... ,α_r) for the proposed SC-MCGM(1,1) model, where w_1, w_2, w_3, w_4, w_5, w_6, and θ range from -1 to 1, and p_1, p_2 ... ,p_{r-1} range from $\varepsilon_{\min}$ to $\varepsilon_{\max}$. For the current population, the best solution was the chromosome with the maximum fitness value.

Let n_{size} and n_{max} denote the population size and maximum number of generations (i.e., GA terminated after n_{max} generations), respectively. Each of the populations consists of n_{size} strings. After evaluating the fitness value of each chromosome in P_m, n_{size} new chromosomes were generated for P_{m+1} by selection, crossover, and mutation. Crossover and mutation reproduce children for a selected parent by changing the parents' chromosomal makeup. When the stopping condition was satisfied, the algorithm terminated, and the best chromosome is the one with maximum fitness value among consecutive generations. This best case can then be used to examine the SC-MCGM(1,1) model generalization.

4.1. Selection

Let P_m denote the population in generation m $(1 \le m \le n_{max})$, where chromosome u $(1 \le u \le n_{size})$ produced in P_m is represented as $w^m_{u,1}\ w^m_{u,2}\ w^m_{u,3}\ w^m_{u,4}\ w^m_{u,5}\ w^m_{u,6}\ \theta^m_u\ p^m_{u,1}\ p^m_{u,2} \ldots\ p^m_{u,r-1}\ \alpha^m_{u,1}\ \alpha^m_{u,2}\ \ldots\ \alpha^m_{u,r}$. Two chromosomes were randomly selected from P_m by binary tournament selection, and the one with higher fitness was put into a mating pool [36]. This process was repeated until n_{size} chromosomes were in the mating pool.

4.2. Crossover

To generate new chromosomes in the next population, $\frac{1}{2}n_{size}$ pairs of chromosomes from the pool were randomly selected from the current population, and offspring of the selected parents were reproduced by crossover and mutation. For chromosomes u and v $(w^m_{v,1}\ w^m_{v,2}\ w^m_{v,3}\ w^m_{v,4}\ w^m_{v,5}\ w^m_{v,6}\ \theta^m_v\ r^m_v\ p^m_{v,1}\ p^m_{v,2} \ldots\ p^m_{v,r-1}\ \alpha^m_{v,1}\ \alpha^m_{v,2}\ \ldots\ \alpha^m_{v,r})$ $(1 \le v \le n_{size})$, each pair of real-valued genes was used to generate two new genes with crossover probability Pr_c,

$$\begin{array}{l}
w^m_{u,i} = h_i w^m_{u,i} + (1-h_i) w^m_{v,i}, w^m_{v,i}\prime = (1-h_i) w^m_{u,i} + h_i w^m_{v,i}, i = 1, 2, \ldots, 6 \\
\theta^m_u\prime = h_7\theta^m_u + (1-h_7)\theta^m_v, \theta^m_v\prime = (1-h_7)\theta^m_u + h_7\theta^m_v \\
p^m_{u,i}\prime = h_{7+j} p^m_{u,i} + (1-h_{7+j}) p^m_{v,i}, p^m_{v,i}\prime = (1-h_{7+j}) p^m_{u,i} + h_{7+j} p^m_{v,i}, j = 1, 2, \ldots, r-1 \\
\alpha^m_{u,i}\prime = h_{6+r+j}\alpha^m_{u,i} + (1-h_{6+r+j})\alpha^m_{v,i}, p^m_{v,i}\prime = (1-h_{6+r+j})\alpha^m_{u,i} + h_{6+r+j}\alpha^m_{v,i}, j = 1, 2, \ldots, r
\end{array} \quad (23)$$

where $h_1, h_2, \ldots, h_{6+2r}$ are random numbers between 0 and 1.

4.3. Mutation

Mutation occurred with probability Pr_m for each real valued gene in a newly generated chromosome produced by crossover. A low mutation rate was used to avoid excessive perturbation. When a mutation occurs, that gene was altered by adding or subtracting a tiny number randomly selected from a pre-specified interval. Subsequently, n_{del} $(0 \le n_{del} \le n_{size})$ chromosomes in P_{m+1} were randomly removed from the set of new chromosomes (formed by genetic operations) to allow additional copies of the chromosome with maximum fitness values in P_m. Only two or three elite chromosomes are sufficient to generate better results [37].

4.4. Algorithm for Constructing the Proposed Model

The GA for constructing the proposed SC-MCGM(1,1) is briefly described below.

Algorithm: GA for constructing the proposed SC-MCGM(1,1).

Step 1. Initialization

Generate n_{size} chromosomes.

Step 2. Compute fitness values

Compute the fitness value of each chromosome in the current population.

Step 3. Generate new chromosomes

Generate n_{size} new chromosomes from the current population using selection, crossover, and mutation.

Step 4. Apply elitist strategy

Randomly remove n_{del} strings from the newly generated n_{size} strings, and replace them with n_{del} best chromosomes in the current population.

Step 5. Termination test

Return to Step 2 if the stopping condition is not satisfied.

5. Empirical Results

Two real data sets were used to compare **foreign tourist** forecasting using the proposed SC-MCGM(1,1) model with different transitions (m = 1,2,3,4) against the original GM(1,1), MCGM(1,1), and several models proposed by Sun et al. [1], including segmented GM(1,1), SGM(1,1) using Markov chain, and MCSGM(1,1) using a Cuckoo search algorithm. They are denoted by SGM(1,1), MCSGM(1,1), and CMCSGM(1,1), respectively.

In contrast to the original GM(1,1) and the MCGM(1,1) models, which use all observed data, the SGM model first used a rolling mechanism to determine the set of newly observed data, and then constructed the GM(1,1) model. The rolling mechanism could select only a few recent data by capturing the developing trend from all observed data. The training data were retained after rolling and applied to the SGM(1,1), MCSGM(1,1), CMCSGM(1,1), and proposed SC-MCGM(1,1) models. Related GA parameters were: (1) population size n_{size} = 200; (2) stopping condition n_{max} = 500; (3) n_{del} = 2; (4) crossover probability Pr_c = 0.9; and (5) mutation probability Pr_m = 0.01.

5.1. Prediction of Foreign Tourists for Taiwan

The first experiment was conducted on the yearly statistics reported by Taiwan Tourism Bureau. Table 1 shows historical annual foreign tourists to Taiwan from six economies, including Japan, Hong Kong/Macao, Korea, China, USA, and Southeast Asia, collected from 2001–2016. Year 2016 was used for ex post testing. After performing the rolling mechanism, 2011–2015 (n = 5) from China and 2012–2015 (n = 4) from the other economies can be used for model-fitting for the SGM(1,1), MCSGM(1,1), CMCSGM(1,1), and proposed SC-MCGM(1,1) models. Both n = 4 and 5 produced two intervals.

Table 1. Historical annual foreign tourists from six economies to Taiwan.

Year	Japan	Hong Kong/Macao	Korea	China	USA	Southeast Asia
2001	976,750	435,164	85,744		348,808	488,968
2002	998,497	456,554	83,624		377,470	530,319
2003	657,053	323,178	92,893		272,858	457,103
2004	887,311	417,087	148,095		382,822	568,269
2005	1,124,334	432,718	182,517		390,929	636,925
2006	1,161,489	431,884	196,260		394,802	643,338
2007	1,166,380	491,437	225,814		397,965	700,287
2008	1,086,691	618,667	252,266	329,204	387,197	725,751
2009	1,000,661	718,806	167,641	972,123	369,258	689,027
2010	1,080,153	794,362	216,901	1,630,735	395,729	911,174
2011	1,294,758	817,944	242,902	1,784,185	412,617	1,071,975
2012	1,432,315	1,016,356	259,089	2,586,428	411,416	1,132,592
2013	1,421,550	1,183,341	351,301	2,874,702	414,060	1,261,596
2014	1,634,790	1,375,770	527,684	3,987,152	458,691	1,388,305
2015	1,627,229	1,513,597	658,757	4,184,102	479,452	1,425,485
2016	1,895,702	1,614,803	884,397	3,511,734	523,888	1,653,908

Table 2 shows prediction results associated with ex post testing for the original GM(1,1), the MCGM(1,1), the SGM(1,1), the MCSGM(1,1), and the CMCSGM(1,1), and Table 3 shows those for the proposed SC-MCGM(1,1) with different values for m. For the proposed SC-MCGM(1,1), it seems that the worse results can be obtained for m = 1. When $m \geq 2$, the results obtained by the proposed SC-MCGM(1,1) is comparable or superior to those described in Table 2. For instance, the SC-MCGM(1,1) with m = 2, 3, 4 outperform the SGM(1,1), MCSGM(1,1), and CMCSGM(1,1) for Japan, Hong Kong/Macao, Korea, China, and Southeast Asia. Since the number of visitors from China to Taiwan dramatically declined in 2016, as shown in Table 1, the results of the ex post testing are relatively poor for every prediction model. However, the proposed SC-MCGM(1,1) outperforms the others, even with this significant temporal fluctuation.

Table 2. Absolute percentage errors obtained by different forecasting methods for Case I.

Economy	GM(1,1)	MCGM(1,1)	SGM(1,1)	MCSGM(1,1)	CMCSGM(1,1)
Japan	11.01	12.41	6.50	5.62	4.27
Hong Kong/Macao	2.40	6.58	6.36	6.83	7.40
Korea	33.35	17.96	0.47	1.21	3.76
China	58.05	61.54	45.79	49.32	50.19
USA	11.40	13.18	0.88	0.45	0.10
Southeast Asia	3.44	1.16	7.59	7.13	6.52

Table 3. Absolute percentage errors obtained by the proposed SC-MCGM(1,1) with different transitions for Case I.

Economy	SC-MCGM(1,1)			
	$m = 1$	$m = 2$	$m = 3$	$m = 4$
Japan	6.41	1.31	3.72	2.96
Hong Kong/Macao	5.52	5.36	4.71	4.60
Korea	3.47	0.38	0.46	0.05
China	40.58	41.60	41.91	40.91
USA	0.78	0.30	0.36	0.59
Southeast Asia	6.94	5.03	4.95	4.78

5.2. Prediction of Foreign Tourists for China

Historical annual data from 1997 to 2013 published by the China National Tourism Administration were used. The collected data described the number of foreign tourists from eight main economies, including Japan, Korea, Malaysia, Mongolia, Philippines, Russia, Singapore, and USA. Following Sun et al. [1], year 2013 was used for ex post testing using a one-step transition probability matrix. Performing the rolling mechanism, 2005–2012 data from Korea, Japan, USA, and Malaysia; 2006–2012 from Russia, 2003–2012 from Mongolia and Philippines; and 2004–2012 from Singapore were used to construct the SGM(1,1), MCSGM(1,1), CMCSGM(1,1), and proposed SC-MCGM(1,1) models.

Table 4 shows prediction results for all compared models, and Table 5 shows those for the proposed SC-MCGM(1,1) with different m. The original GM(1,1) and MCGM(1,1) were significantly poorer than the other methods considered. The results obtained by the proposed SC-MCGM(1,1) with different m is comparable or superior to those described in Table 4. For instance, the SC-MCGM(1,1) outperform the prediction methods considered for Korea and Mongolia for all m considered, and outperform the compared models for USA and Singapore for m = 2, 3, 4.

Table 4. Absolute percentage errors obtained by different forecasting methods for Case II.

Economy	GM(1,1)	MCGM(1,1)	SGM(1,1)	MCSGM(1,1)	CMCSGM(1,1)
Korea	35.90	30.38	0.84	1.32	0.97
Japan	39.57	15.26	20.93	21.02	15.69
Russia	44.13	36.42	2.72	3.05	2.90
USA	20.5	13.68	5.22	2.95	2.92
Malaysia	28.81	27.81	11.73	11.50	7.24
Mongolia	1.39	3.12	0.79	1.62	1.82
Philippines	10.16	7.63	0.67	0.69	0.91
Singapore	31.68	18.92	12.85	5.83	8.03

Table 5. Absolute percentage errors obtained by the proposed SC-MCGM(1,1) with different transitions for Case II.

Economy	SC-MCGM(1,1)			
	$m = 1$	$m = 2$	$m = 3$	$m = 4$
Korea	0.47	0.72	0.77	0.77
Japan	21.08	20.74	20.17	20.83
Russia	1.01	1.63	0.67	3.40
USA	5.36	0.90	2.38	1.36
Malaysia	10.79	9.86	7.03	7.62
Mongolia	0.16	0.11	0.27	0.23
Philippines	0.23	0.07	1.66	1.00
Singapore	11.04	5.22	6.74	6.54

5.3. Statistical Analysis

We investigated if the SC-MCGM(1,1) model prediction accuracy could be improved by tuning m, compared to the $m = 1$ case. We compared outcomes for the fourteen data sets (six economies for Taiwan and eight for China) using the non-parametric Friedman test [38] with the post-hoc test, the Bonferroni–Dunn test [39]. The Friedman test checks whether average ranks are significantly different, ranking the SC-MCGM(1,1) model for all m separately for each data set, with increasing rank number implying lower prediction accuracy. In case of ties, the average rank was assigned.

Let r_j, k_1, and k_2 denote the average rank of prediction method j, number of prediction methods, and number of data sets used, respectively; and prediction methods 1, 2, 3, and 4 denote SC-MCGM(1,1) with m = 1, 2 , 3, and 4, respectively. Therefore, $k_1 = 4$, $k_2 = 14$; and r_1 = 3.14, 1.93, 2.61, and 2.32 for m = 1, 2, 3, and 4, respectively. The Friedman statistic [40],

$$F_F = \frac{(k_2 - 1)\chi_F^2}{k_2(k_1 - 1) - \chi_F^2} \tag{24}$$

is distributed according to the F distribution, and χ_F^2 is

$$\chi_F^2 = \frac{12k_2}{k_1(k_1+1)}\left[\sum_{j=1}^{k_1} r_j^2 - \frac{k_1(k_1+1)^2}{4}\right] \tag{25}$$

Since $F_F = 2.41$ is greater than the critical value $F(k_1 - 1, (k_1 - 1)(k_2 - 1))$ at the 10% level (i.e., $F(3, 39) = 2.23$), the null hypothesis is rejected.

Subsequently, the Bonferroni–Dunn test was used to detect any significant differences among m = 1, 2, 3, and 4. Significant differences in the prediction accuracy of the two forecasting methods can be probed by the difference in their average ranks, with the critical difference at the 10% level being

$$CD = q_{0.1}\sqrt{\frac{k_1(k_1+1)}{6k_2}} \tag{26}$$

where $q_{0.10} = 2.128$. The difference is significant when $CD > 1.03$. Thus, SC-MCGM(1,1) with $m = 2$ is significantly superior than with $m = 1$, whereas $m = 3$ and 4 are not. Thus, long term transitions in the Markov model could be useless for tourism demand forecasting.

6. Discussion and Conclusions

Accurate foreign tourist forecasting is critical for governmental tourism development policies. However, time series data for tourism often have temporal fluctuation and trend changes, making precise predictions challenging. Over or under estimation of foreign tourist numbers could lead to inappropriate governmental investment in tourist infrastructures [41]. This study proposed a novel grey residual modification model incorporating soft computing, including SLP, FLN, and GA, into the Grey–Markov model.

Historical annual data for foreign tourists collected from Taiwan and China official institutions were used to evaluate prediction accuracy of the proposed model. Soft computing constructs a computationally intelligent system that can learn to achieve better accuracy for changing environments and confront real world problems [42]. Therefore, this study proposed a residual modification model called SC-MCGM(1,1) by incorporating soft computing techniques into the Grey–Markov model. The proposed SC-MCGM(1,1) model was capable of estimating not only the adjusted volume associated with a new residual value for a predicted value from the GM(1,1), but also avoided the troublesome bounds of individual states for the Markov chain. The degree to which a representative point in each state can contribute to a new residual value was based on the sum of transition probability matrices with different steps.

It is found that the growth rate of foreign tourists to Taiwan from Japan and Korea is greatly increased to 22% in 2016. To effectively stimulate the increase of the number of inbound visitors from Northeast Asia, Taiwan authorities should think about how to further work in close cooperation with related economies. Several relevant tourism development and marketing strategies should be put forth to promote the tourism industry. For instance, more bilateral routes that should be taken into account, the promotion for more attractive itineraries with shopping and accommodation, and environmental impact assessment. Since the SC-MCGM(1,1) model among the compared models performs very well for predicting visitors from Northeast Asia, this suggests that Taiwan authorities can leverage the proposed model to set up tourism development plans for Northeast Asia for a few years. As for China, the SC-MCGM(1,1) model performs very well for predicting visitors from Korea, Russia, US, Mongolia, Philippines, and Singapore. Similarly, China authorities can set up appropriate tourism policies by using the proposed SC-MCGM(1,1) model to predict the number of inbound visitors from those economies for a few years. After all, a prediction model can play a significant role on implementation of tourism development plans [41].

To maximize prediction accuracy, representative points, state bounds for the Markov chain, and FLN connection weights were automatically determined by GA, using a relatively simple computer program. Historical annual data involving foreign tourist collected from Taiwan and China official institutions were used to evaluate prediction performance of the proposed SC-MCGM(1,1). As for ex post testing, we can see that the proposed SC-MCGM(1,1) with $m = 2$ is comparable or superior to the compared models. Tables 1 and 3 show that the proposed model outperforms the other methods for 11 of the 14 data sets. These results validate that combining neural networks and GA is advantageous for intelligent prediction models.

The proposed SC-MCGM(1,1) with pre-specified GA parameters, including population size, number of generations, and crossover and mutation parameters, performed well. Thus, fine parameter tuning was not required. FLN used the hyperbolic tangent as the output neuron's activation function, computing a weighted sum of a vector of connection weights with an enhanced pattern. This assumes additivity among individual variables in the enhanced pattern [43]. However, these criteria are not always independent [9,43–48]. Therefore, future research will investigate the impact of non-additivity on prediction performance of the proposed SC-MCGM(1,1) model.

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 declares no conflict of interest.

References

1. Sun, X.; Sun, W.; Wang, J.; Gao, Y. Using a Grey-Markov model optimized by Cuckoo search algorithm to forecast the annual foreign tourist arrivals to China. *Tour. Manag.* **2016**, *52*, 369–379. [CrossRef]

2. Deng, J.L. Control problems of grey systems. *Syst. Control Lett.* **1982**, *1*, 288–294.
3. Liu, S.; Lin, Y. *Grey Information: Theory and Practical Applications*; Springer: Berlin, Germany, 2010.
4. Liu, S.; Yang, Y.; Forrest, J. *Grey Data Analysis: Methods, Models and Applications*; Springer: Berlin, Germany, 2017.
5. Suganthi, L.; Samuel, A.A. Energy models for demand forecasting—A review. *Renew. Sust. Energ. Rev.* **2012**, *16*, 1223–1240. [CrossRef]
6. Cui, J.; Liu, S.F.; Zeng, B.; Xie, N.M. A novel grey forecasting model and its optimization. *Appl. Math. Model.* **2013**, *37*, 4399–4406. [CrossRef]
7. Feng, S.J.; Ma, Y.D.; Song, Z.L.; Ying, J. Forecasting the energy consumption of China by the grey prediction model. *Energ. Source Part B* **2012**, *7*, 376–389. [CrossRef]
8. Hsu, C.C.; Chen, C.Y. Applications of improved grey prediction model for power demand forecasting. *Energy Convers. Manag.* **2003**, *44*, 2241–2249.
9. 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.
10. 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]
11. 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]
12. Mao, M.Z.; Chirwa, E.C. Application of grey model GM(1,1) to vehicle fatality risk estimation. *Technol. Forecast. Soc. Chang.* **2006**, *73*, 588–605. [CrossRef]
13. Wei, J.; Zhou, L.; Wang, F.; Wu, D. Work safety evaluation in Mainland China using grey theory. *Appl. Math. Model.* **2015**, *39*, 924–933. [CrossRef]
14. Hu, Y.C.; Jiang, P. Forecasting energy demand using neural-network-based grey residual modification models. *J. Oper. Res. Soc.* **2017**, *68*, 556–565. [CrossRef]
15. He, Y.; Bao, Y.D. Grey-Markov forecasting model and its application. *Syst. Eng.-Theory Pract.* **1992**, *9*, 59–63.
16. Hsu, Y.T.; Liu, M.C.; Yeh, J.; Hung, H.F. Forecasting the turning time of stock market based on Markov-Fourier grey model. *Expert Syst. Appl.* **2009**, *36*, 8597–8603. [CrossRef]
17. Wang, C.N.; Phan, V.T. An improved nonlinear grey Bernoulli model combined with Fourier series. *Math. Probl. Eng.* **2015**. [CrossRef]
18. Kumar, U.; Jain, V.K. Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India. *Energy* **2010**, *35*, 1709–1716. [CrossRef]
19. Li, G.D.; Masuda, S.; Nagai, M. The prediction model for electrical power system using an improved hybrid optimization model. *Int. J. Electr.* **2013**, *44*, 981–987. [CrossRef]
20. Xie, N.M.; Yuan, C.Q.; Yang, Y.J. Forecasting China's energy demand and self-sufficiency rate by grey forecasting model and Markov model. *Int. J. Electr.* **2015**, *66*, 1–8. [CrossRef]
21. Hsu, C.I.; Wen, Y.U. Improved Grey prediction models for trans-Pacific air passenger market. *Transport Plan. Technol.* **1998**, *22*, 87–107. [CrossRef]
22. Hsu, L.C. Applying the grey prediction model to the global integrated circuit industry. *Technol. Forecast Soc. Chang.* **2003**, *70*, 563–574. [CrossRef]
23. Mao, Z.L.; Sun, J.H. Application of Grey-Markov model in forecasting fire accidents. *Procedia Eng.* **2011**, *11*, 314–318.
24. Wang, C.H. Predicting tourism demand using fuzzy time-series and hybrid grey theory. *Tour. Manag.* **2004**, *25*, 367–374. [CrossRef]
25. Goldberg, D.E. *Genetic Algorithms in Search, Optimization, and Machine Learning*; Addison-Wesley: Boston, MA, USA, 1989.
26. Ishibuchi, H.; Nakashima, T.; Nii, M. *Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining*; Springer: Heidelberg, Germany, 2004.
27. Kuncheva, L.I. *Fuzzy Classifier Design*; Physica-Verlag: Heidelberg, Germany, 2000.
28. Osyczka, A. *Evolutionary Algorithms for Single and Multicriteria Design Optimization*; Physica-Verlag: Heidelberg, Germany, 2003.
29. Hu, Y.C. Functional-link nets with genetic-algorithm-based learning for robust nonlinear interval regression analysis. *Neurocomputing* **2009**, *72*, 1808–1816. [CrossRef]
30. Pao, Y.H. *Adaptive Pattern Recognition and Neural Networks*; Addison-Wesley: Boston, MA, USA, 1989.

31. Pao, Y.H. Functional-link net computing: Theory, system architecture, and functionalities. *Computer* **1992**, *25*, 76–79. [CrossRef]
32. 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]
33. Hu, Y.C. Electricity consumption forecasting using a neural-network-based grey prediction approach. *J. Oper. Res. Soc.* **2016**. [CrossRef]
34. Imbusch, G.F.; Yen, W.M. The McCumber and sturge formula. *J. Lumin.* **2000**, *85*, 177–179. [CrossRef]
35. 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]
36. Hu, Y.C. A multicriteria collaborative filtering approach using the indifference relation and its application to initiator recommendation for group-buying. *Appl. Artif. Intell.* **2014**, *28*, 992–1008. [CrossRef]
37. Murata, T.; Ishibuchi, H.; Tanaka, H. Multi-objective genetic algorithm and its applications to flowshop scheduling. *Comput. Ind. Eng.* **1989**, *30*, 957–968. [CrossRef]
38. Friedman, M. A comparison of alternative tests of significance for the problem of m rankings. *Ann. Math. Stat.* **1940**, *11*, 86–92. [CrossRef]
39. Demšar, J. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.
40. Iman, R.L.; Davenport, J.M. Approximations of the critical region of the Friedman statistic. *Commun. Stat.* **1980**, *9*, 571–595. [CrossRef]
41. Lin, C.J.; Chen, H.F.; Lee, T.S. Forecasting tourism demand using time series, artificial neural networks and multivariate adaptive regression splines: Evidence from Taiwan. *Int. J. Bus. Adm.* **2011**, *2*, 14–24.
42. Jang, J.S.R; Sun, C.T.; Mizutani, E. *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*; Prentice-Hall, Upper Saddle River, NJ, USA, 1997.
43. Hu, Y.C.; Tseng, F.M. Functional-link net with fuzzy integral for bankruptcy prediction. *Neurocomputing* **2007**, *70*, 2959–2968. [CrossRef]
44. Hu, Y.C. Nonadditive grey single-layer perceptron with Choquet integral for pattern classification problems using genetic algorithms. *Neurocomputing* **2008**, *72*, 332–341. [CrossRef]
45. Onisawa, T.; Sugeno, M.; Nishiwaki, M.Y.; Kawai, H.; Harima, Y. Fuzzy measure analysis of public attitude towards the use of nuclear energy. *Fuzzy Set Syst.* **1986**, *20*, 259–289. [CrossRef]
46. Wang, Z.; Leung, K.S.; Klir, G.J. Applying fuzzy measures and nonlinear integrals in data mining. *Fuzzy Set. Syst.* **2005**, *156*, 371–380. [CrossRef]
47. Wang, Z.; Leung, K.S.; Wang, J. A genetic algorithm for determining nonadditive set functions in information fusion. *Fuzzy Set. Syst.* **1999**, *102*, 463–469. [CrossRef]
48. Wang, W.; Wang, Z.; Klir, G.J. Genetic algorithms for determining fuzzy measures from data. *J. Intell. Fuzzy Syst.* **1998**, *6*, 171–183.