

Research Article

An Improved Metabolism Grey Model for Predicting Small Samples with a Singular Datum and Its Application to Sulfur Dioxide Emissions in China

Wei Zhou and Demei Zhang

International Business School, Yunnan University of Finance and Economics, Kunming 650221, China

Correspondence should be addressed to Wei Zhou; zhouwei907@ynufe.edu.cn

Received 15 December 2015; Revised 24 January 2016; Accepted 27 January 2016

Academic Editor: Filippo Cacace

Copyright © 2016 W. Zhou and D. Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study proposes an improved metabolism grey model [IMG $M(1, 1)$] to predict small samples with a singular datum, which is a common phenomenon in daily economic data. This new model combines the fitting advantage of the conventional GM(1, 1) in small samples and the additional advantages of the MGM(1, 1) in new real-time data, while overcoming the limitations of both the conventional GM(1, 1) and MGM(1, 1) when the predicted results are vulnerable at any singular datum. Thus, this model can be classified as an improved grey prediction model. Its improvements are illustrated through a case study of sulfur dioxide emissions in China from 2007 to 2013 with a singular datum in 2011. Some features of this model are presented based on the error analysis in the case study. Results suggest that if action is not taken immediately, sulfur dioxide emissions in 2016 will surpass the standard level required by the Twelfth Five-Year Plan proposed by the China State Council.

1. Introduction

In response to technological advances and the progress of human society, many scholars have proposed and investigated various theories and methods that analyze uncertain information from different angles and perspectives. The grey system theory as proposed by Chinese scholar Julong Deng in 1982 is a novel uncertainty theory, which has received increasing attention and has recently become a popular subject of research. The grey system is an uncertainty system in which information is only partially known; it is also suitable to modeling small samples and poor information.

The grey prediction models, especially the GM(1, 1), may be superior to other prediction techniques in the context of small samples and poor information. Thus, grey prediction models have received much attention since their introduction and successful application in the small sample prediction fields. They have been applied in many fields. Wang [1] predicted the stock price using a fuzzy grey prediction system. Sun et al. [2] proposed a small sample prediction technology for early warning of financial crises and recognition of crisis

patterns of large-scale enterprises based on the GM(1, 1). Vishnu and Syamala [3] used the basic GM(1, 1) to predict dynamic changes in water supply systems. Chang and Lu [4] used the GM(1, 1) predictor to forecast the flicker severity. Li et al. [5] used the adaptive GM(1, 1) to forecast short-term electricity consumption. Tang and Yin [6] forecasted the education expenditure and school enrollment by the GM(1, 1). Yin and Tang [7] fitted and forecasted China's labor formation based on the prediction GM(1, 1). Li et al. [8] applied the GM(1, 1) and GRA to evaluate the financial burden of patients at hospitals in China by PPP model. Wu et al. [9] applied an improved GM(1, 1) to analyze the trends of the plum virus incidence in China. Tabaszewskin and Cempel [10] developed a grey prediction method for the fan vibration in a cogeneration plant (CHP) by combining the GM(1, 1) and the optimal estimated approaches. Pao et al. [11] forecasted CO₂ emissions, energy consumption, and economic growth in China using a prediction entirely improved GM(1, 1). Recently, Huang et al. [12] constructed a hybrid nonlinear grey prediction model to analyze the carbon reduction at dual levels in China. Wang et al. [13] provided

an approach to forecast the coal mine water inflow based on the GM(1, 1) prediction theory. Ma and Fu [14] studied the accidents due to human factor in coal mine based on the GM(1, 1). These applications of the grey prediction models show the feasibility and superiority of the grey system theory in dealing with small samples, which is also a reason why we choose this method to predict sulfur dioxide emissions.

With respect to the limitations of the basic grey prediction models, many scholars have proposed some optimization GM(1, 1). Xie and Liu [15, 16] proposed and studied a discrete GM(1, 1) and verified its unbiasedness to fit an exponential sequence. Yao and Liu [17] and Wang et al. [18], respectively, developed an improved and optimal discrete GM(1, 1) based on Xie's discrete GM(1, 1). Shih et al. [19] improved the GM(1, 1) by changing the background and initial values. Li et al. [20, 21] provided an overall definition of the grey prediction model and proposed some basic GM(1, 1). Truong and Ahn [22] proposed the SAGM(1, 1) to overcome the instability and improve the prediction accuracy of the GM(1, 1). Cui et al. [23] provided the NGM(1, 1) based on the GM(1, 1), which belongs to an entirely optimized GM(1, 1). Dai and Huang [24] developed the GM(1, 1) by designing calculation program and selecting the initial value. Zhou and He [25] proposed the GGM(1, 1) which is generalized GM(1, 1) that includes the DGM, SAGM, and NGM. Lin and Lian [26] designed a grey prediction self-organizing model, which is an extended GM(1, 1). Xiong et al. [27] proposed the nonequidistant GM(1, 1) based on the optimization initial condition. The main objectives of the above methods are to improve the accuracy of grey prediction models and, thereby, perfect them. However, these methods have a common data background referred to as small samples, which demonstrate an increasing or decreasing trend. This simple sequence does not include a small simple sequence with several singular data or a nonmonotone sequence, which generally constitute a practical phenomenon in the periodicity of economic and environmental uncertainty. For this reason, this study investigated the applicability of a new grey prediction model in a small simple sequence with a singular datum, which is a key point in the prediction of sulfur dioxide emissions in China.

Due to the recently increasing interest in environmental protection, many studies have focused on predicting energy consumption and pollution emissions. However, the prediction results of such studies remain relatively inaccurate due to changing environmental indicators and nonuniform standards. To solve this problem, some scholars have successfully applied the grey system theory and its prediction models to predict the environmental indicators with small samples. For example, Wang et al. [28] applied the GM(1, 1) to analyze and predict the atmospheric ozone levels in Asia. Lee and Tong [29] proposed the GPGM(1, 1) to forecast the energy consumption in China and achieved significant fitting results. Lin et al. [30] used the GM(1, 1) to predict the carbon dioxide emissions in China. Pan et al. [31] and Dai et al. [32] applied the GM(1, 1) to predict the degrees of air pollution in Tianjin and Shenzhen, respectively. These studies prove that the grey prediction model is suitable and useful in forecasting the energy consumption and pollution emission

in China. However, the grey prediction model has yet to be investigated with small samples of singular data, which have become increasingly common in Chinese economic data to reflect the decline in the economic growth.

Based on the literature review above and practical trends in Chinese pollution indicators, this study proposes a new grey model to predict small samples with singular data and applies it to predict recent annual sulfur dioxide emissions in China. This paper is organized as follows: Section 2 describes the basic GM(1, 1) and MGM(1, 1). Section 3 analyzes the limitations of the GM(1, 1) and MGM(1, 1) and then proposes the improved MGM(1, 1) called IMG(1, 1), based on which it provides the corresponding modeling steps. In Section 4, a practical example is provided to demonstrate the effectiveness and the practicability of the new grey prediction model. The paper ends in Section 5 with conclusions.

2. The GM(1, 1) and MGM(1, 1)

2.1. The Modeling Mechanism of the GM(1, 1). Grey system theory, the name of which was derived from cybernetics and a clear degree of information, was proposed by Chinese scholar Julong Deng in 1982 [33]. In economics management processes and scientific research, incomplete information can arise in a general situation belonging to a grey system, as defined in grey system theory. As such, a grey prediction model has a comprehensive scope of applications in our society. However, a prerequisite for the application of grey prediction models, such as the GM(1, 1), is that the original data must be monotonically increasing or decreasing. If this prerequisite is not met, the prediction results are likely to seriously deviate from the reality. This prerequisite and its influence are investigated through the following analysis of the modeling mechanism of the GM(1, 1).

Let a nonnegative small sample sequence $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$ be the original data sequence, $x^{(0)}(k) \geq 0$ and $k = 1, 2, \dots, n$. As a small sample sequence, n is generally small at $n \leq 10$. Deng [33] proposed the GM(1, 1) and its modeling steps to fit and predict based on $X^{(0)}$ as follows.

Step 1. Use the one-time accumulated generating operation (1-AGO) to obtain $X^{(1)}$, in which $X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$ and

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), \quad k = 1, 2, \dots, n. \quad (1)$$

Step 2. Calculate the background value $Z^{(1)}$, in which $Z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(k)), k = 2, 3, \dots, n$, and

$$z^{(1)}(k) = \frac{1}{2} (x^{(1)}(k) + x^{(1)}(k-1)). \quad (2)$$

Step 3. Construct the original form of GM(1, 1) $X^{(0)}(k) + aZ^{(1)}(k) = b$, and then estimate the developing coefficient $-a$ and the grey action quantity b based on the ordinary least squares method presented as follows:

$$\alpha = (a, b)^T = (B^T B)^{-1} B^T Y, \quad (3)$$

where $\hat{a} = (a, b)^T$, $Y = [x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(k)]^T$, and

$$B = \begin{bmatrix} -z^{(1)}(2), -z^{(1)}(3), \dots, -z^{(1)}(n) \\ 1, 1, \dots, 1 \end{bmatrix}^T. \quad (4)$$

Step 4. Calculate the time response sequence $\bar{x}^{(1)}(k+1)$ through the following formula:

$$\bar{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a}. \quad (5)$$

Step 5. The time response sequence $\bar{x}^{(1)}(k+1)$ is transformed into the restored time response sequence $\bar{x}^{(0)}(k+1)$ through the following formula:

$$\begin{aligned} \bar{x}^{(0)}(k+1) &= \bar{x}^{(1)}(k+1) - \bar{x}^{(1)}(k) \\ &= (1 - e^a) \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ak}. \end{aligned} \quad (6)$$

The abovementioned steps provide the modeling process of the GM(1, 1), which also is its modeling mechanism. Here, $\bar{x}^{(1)}(n+j)$ is the cumulative prediction value, $\bar{x}^{(0)}(n+j)$ is the restored prediction value, $\bar{x}^{(1)}(k)$ is the cumulative fitting value, and $\bar{x}^{(0)}(k)$ is the restored fitting value, where $k = 1, 2, \dots, n$ and $j = 1, 2, \dots$

With respect to the above process, the GM(1, 1) was found to have two advantages and two shortcomings. The *advantages* are as follows: (1) The model does not require many samples and is simpler and more suitable than general econometric models. (2) It can be used for recent or short-term prediction activities. The *shortcomings* are as follows: (1) The model is a monotonic exponential prediction approach, which is its modeling prerequisite. (2) It models and predicts all data when $t = n$, but it ignores new information and it cannot accurately reflect the characteristics of the current situation.

Liu et al. [34] proposed the metabolism grey model, namely, MGM(1, 1), by updating the modeling data and introducing new information to overcome the second shortcoming, which is a significant improvement as explained in Section 2.2. In Section 3, we propose and analyze another new GM(1, 1) based on the above MGM(1, 1) to overcome the first shortcoming.

2.2. Construction and Calculation of the MGM(1, 1). As previously stated, the general grey prediction models, such as the GM(1, 1), provide forecasting based on the original data without considering the impact of new information or dynamic predictions when information is subsequently supplied and updated. As such, Liu et al. [34] proposed the MGM(1, 1). As an improved grey prediction model, the MGM(1, 1) can add new information and remove old information in what is also a metabolic process. The corresponding modeling steps of the MGM(1, 1) are presented as follows.

Step 1. Construct the GM(1, 1) for a small sample sequence also called original data sequence, and then calculate the time

response sequence $\bar{x}^{(1)}(k+1)$ and the restored time response sequence $\bar{x}^{(0)}(k+1)$ based on (5) and (6), where the original data sequence is $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$ and $k = 1, 2, \dots, n$.

Step 2. Eliminate and metabolize the fitting data. In particular, the new prediction data $\bar{x}^{(0)}(n+1)$ is used instead of the first data $x^{(0)}(1)$. Thus, we can obtain a metabolic small sample sequence $X^{(0)} = (x^{(0)}(2), x^{(0)}(3), \dots, \bar{x}^{(0)}(n+1))$ called metabolic data sequence, which is similar to the original data sequence and includes a new datum.

Step 3. Use the MGM(1, 1) to calculate and predict, following the same modeling process of the GM(1, 1). Then, a new prediction datum $\bar{x}^{(0)}(n+2)$ can be obtained.

If more prediction data is needed, then the above process is repeated to obtain the data. This calculation process demonstrates the modeling steps of the MGM(1, 1).

Unlike the GM(1, 1), the MGM(1, 1) uses the prediction data as the original data to predict other data. Therefore, there could be a propagation phenomenon of the prediction error from the MGM(1, 1), which could further affect the metabolic modeling process. Thus, there are two error inspection methods (Step 4) to improve the MGM(1, 1).

Step 4. The fitting errors are calculated based on the following equations.

- (1) The residual error $e^{(0)}(t)$ and its corresponding form are as follows:

$$e^{(0)}(t) = x^{(0)}(t) - \bar{x}^{(0)}(t). \quad (7)$$

- (2) The relative error $e(t)$ and its corresponding form are as follows:

$$e(t) = \frac{x^{(0)}(t) - \bar{x}^{(0)}(t)}{x^{(0)}(t)}. \quad (8)$$

According to the above calculation process of the MGM(1, 1), this new grey prediction model is a small sample and updated information prediction model. Therefore, the MGM(1, 1) is more suitable than the GM(1, 1) in the prediction of changing small simple sequences.

3. The IMGGM(1, 1)

3.1. Shortage Analysis of the GM(1, 1) and MGM(1, 1). Based on the modeling processes of the GM(1, 1) and MGM(1, 1) and the analysis of their disadvantages in the previous section, the original data sequence, which is a monotonic exponential sequence for these two grey prediction models, is taken as a prerequisite. Therefore, if the monotonic exponential character is changed or if singular data exist in the original data, then the prediction results can be affected, and a larger error could result, which can be proved by (9) and the following analysis:

$$\frac{\bar{x}^{(0)}(k+1)}{\bar{x}^{(0)}(k)} = \frac{(1 - e^a) \left(x^{(0)}(1) - b/a \right) e^{-ak}}{(1 - e^a) \left(x^{(0)}(1) - b/a \right) e^{-a(k-1)}} = \frac{1}{e^a}. \quad (9)$$

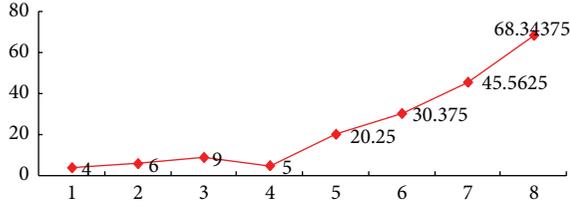


FIGURE 1: Original data sequence of Example 1.

Equation (9) exhibits the following conclusions:

- (i) Given that the ratio between $\bar{x}^{(0)}(k+1)$ and $\bar{x}^{(0)}(k)$ in the GM(1, 1) is a constant and is related to the development coefficient $-a$, a monotonic exponential trend exists in this grey prediction model.
- (ii) This monotonic exponential trend also exists in the MGM(1, 1), which leads to its prerequisite of the monotonic exponential sequence. In our opinion, this prerequisite is unsuitable.

We provide a simple example below to further show the conclusions and limitations discussed above.

Example 1. Let a small sample sequence with a singular datum be

$$X_1^{(0)} = (4, 6, 9, 5, 20.25, 30.375, 45.5625, 68.34375). \quad (10)$$

The distribution and trend of this sequence are shown in Figure 1. This sequence is similar to a monotonic exponential sequence, but it has a singular datum "5."

Let the first five numbers be the original data sequence, that is, $X^{(0)} = (4, 6, 9, 5, 20.25)$, and let the last three numbers be the future and unknown data sequence for comparing with the prediction results. It is found that this original data sequence almost meets two prerequisites of the GM(1, 1) and MGM(1, 1), namely, the small simple and the increasing trend. Thus, the modeling processes of these two models can be used, and then the corresponding results can be obtained, which are shown as follows.

Step 1. Use the 1-AGO operator to obtain

$$X^{(1)} = (4, 10, 19, 24, 44.25). \quad (11)$$

Step 2. Calculate the background values as follows:

$$Z^{(1)} = (7, 14.5, 21.5, 34.125). \quad (12)$$

Step 3. Estimate the developing coefficient $-a$ and the grey action quantity b through the ordinary least squares method and the results shown as follows:

$$B = \begin{bmatrix} -7, & -14.5, & -21.5, & -34.125 \\ 1, & 1, & 1, & 1 \end{bmatrix}^T, \quad (13)$$

$$Y = [6, 9, 5, 20.25]^T,$$

$$\alpha = (a, b)^T = (B^T B)^{-1} B^T Y$$

$$= [-0.488686, 0.640021].$$

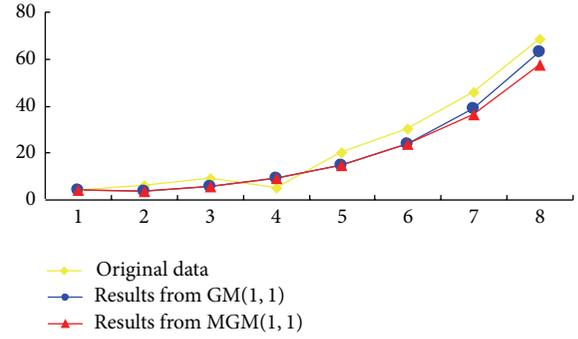


FIGURE 2: Comparison of the three sequences.

Step 4. Calculate the time response sequence $\bar{x}^{(1)}(k+1)$ shown as follows:

$$\bar{x}^{(1)}(k+1) = 5.309677e^{0.488686k} - 1.309677. \quad (14)$$

Step 5. Transform the time response sequence $\bar{x}^{(1)}(k+1)$ into the restored time response sequence $\bar{x}^{(0)}(k+1)$ as follows:

$$\begin{aligned} \bar{x}^{(0)}(k+1) &= \bar{x}^{(1)}(k+1) - \bar{x}^{(1)}(k) \\ &= (1 - 5.309677e^{-0.488686}) 5.309677e^{0.488686 \times k}. \end{aligned} \quad (15)$$

Thus, the modeling steps of the GM(1, 1) are completed. Let $k = 1, 2, 3, 4, 5$; the fitting values of the five original numbers can be obtained. Let $k = 6, 7, 8$; the three prediction values based on the five original numbers can be obtained. The results and the corresponding fitting and prediction errors are shown in Table 1.

We also provide the following fitting and prediction results using the MGM(1, 1). In the calculation process of this model, the original data still follow $X^{(0)} = (4, 6, 9, 5, 20.25)$. After two metabolism computations, the results and corresponding fitting and prediction errors are obtained shown in Table 2.

In Figure 2, the three sequences include the original data sequence, the fitting and prediction result sequence of the GM(1, 1), and the fitting and prediction result sequence of the MGM(1, 1).

Based on Tables 1 and 2 and Figure 2, the following conclusions can be derived.

- (i) A clear monotonic exponential trend is observed in the results of the GM(1, 1) and MGM(1, 1), which is consistent with previous analysis.
- (ii) The fourth data point is a singular datum in the original data, and it cannot be shown in the fitting results of the GM(1, 1) and MGM(1, 1).
- (iii) The GM(1, 1) and MGM(1, 1) can fit three data points before the singular datum.
- (iv) The fitting results after the singular datum and the prediction results are inappropriate for the above second conclusion. The later prediction results are worse than the initial results.

TABLE 1: Fitting and prediction results of the GM(1, 1).

Original data sequence	Modeling data sequence	Fitting and prediction values	Residual errors	Relative errors (%)
4	4	4	2.6539	0
6	6	3.3460	3.5454	44.2330
9	9	5.4545	3.8919	39.3935
5	5	8.8919	5.7546	77.8382
20.25	20.25	14.495	6.7450	28.4180
30.375	—	23.6299	7.0416	22.2059
45.5625	—	38.5208	5.5480	15.4548
68.34375	—	62.7957	2.6539	8.1178

TABLE 2: Fitting and prediction results of the MGM(1, 1).

Original data sequence	Fitted and predictive value	Residual errors	Relative errors
4	4	2.6539	0
6	3.3460	3.5454	44.2330
9	5.4545	3.8919	39.3935
5	8.8919	5.7546	77.8382
20.25	14.4953	6.7450	28.4180
30.375	23.6299	7.0416	22.2059
45.5625	36.2875	9.2749	20.3565
68.34375	57.3535	10.9902	16.0808

Therefore, the GM(1, 1) and MGM(1, 1) are unsuitable to fit and predict the small simple sequences with singular data. Because of this, we propose the IMGGM(1, 1) in Section 3.2.

3.2. *IMGGM(1, 1)*. In this subsection, the IMGGM(1, 1) is proposed by combining the GM(1, 1) and MGM(1, 1), which can be used to effectively fit and predict the small simple sequence with a singular datum. This new grey prediction model combines the fitting advantages of the GM(1, 1) in the small sample sequences and the updated advantages of the MGM(1, 1) in new real-time data. It also overcomes their limitations when the predicted results are vulnerable to any singular datum. Thus, this IMGGM(1, 1) belongs to an improved grey prediction model, and its modeling steps are summarized as follows.

Let a small sample sequence with a singular datum be

$$X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(k^*), \dots, x^{(0)}(n)), \quad (16)$$

where $x^{(0)}(k^*)$ is the singular datum.

Step 1. Construct and apply the GM(1, 1), and the corresponding calculation steps can be referenced in Section 2.1. Thus, we can obtain the time response sequence $\bar{x}_1^{(1)}(k+1)$ and the restored time response sequence $\bar{x}_1^{(0)}(k+1)$, which showed in (5) and (6), respectively, where $k = 1, 2, \dots, n$. Furthermore, the fitting value of the singular datum $x^{(0)}(k^*)$ can be shown as follows:

$$\bar{x}_1^{(0)}(k^*) = (1 - e^a) \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ak^* - 1}. \quad (17)$$

Step 2. Estimate the singular datum, and then update it by substituting $x^{(0)}(k^*)$ with its first fitting value $\bar{x}_1^{(0)}(k^*)$, and a new original sequence is obtained as follows:

$$X^{t(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, \bar{x}_1^{(0)}(k^*), \dots, x^{(0)}(n)). \quad (18)$$

Step 3. Construct and apply the GM(1, 1) again based on the new original sequence $X^{t(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, \bar{x}_1^{(0)}(k^*), \dots, x^{(0)}(n))$. Then we can obtain the second-time time response sequence $\bar{x}_2^{(1)}(k+1)$ and the second-time restored time response sequence $\bar{x}_2^{(0)}(k+1)$, which are different from $\bar{x}_1^{(1)}(k+1)$ and $\bar{x}_1^{(0)}(k+1)$.

Step 4. Derive the metabolism process for the original sequence which can be referenced in the modeling process of the MGM(1, 1), and then construct and apply the GM(1, 1) again based on two new sequences.

Step 5. Calculate the fitting errors based on the above error equations.

Thus, the modeling process of the IMGGM(1, 1) is completed. Step 2 can effectively eliminate the singular datum and provide a monotonic exponential sequence to construct grey prediction models. Step 4 is a metabolic process similar to the MGM(1, 1). Steps 1 and 3 are the modeling process of the GM(1, 1). Step 5 is unique to the IMGGM(1, 1). The new grey prediction model combines the GM(1, 1) and MGM(1, 1) and eliminates the inappropriate influence of singular datum.

TABLE 3: Sulfur dioxide emissions of China from 2007 to 2013.

Year	2007	2008	2009	2010	2011	2012	2013
Emission data (unit: 10^4 tons)	2468.1	2321.2	2214.4	2185.1	2217.9	2117.6	2044

Therefore, we think that the IMGGM(1, 1) is a new and improved grey prediction model that can be used to fit and predict the small simple sequence with a singular datum. This new model deals with a singular datum only, which is a limitation. Although likely to be a challenging undertaking, future studies will conduct a more general situation than a small simple sequence with singular data.

4. Empirical Application and Analysis

4.1. Background. Sulfur dioxide is an environmental pollutant that is a major component of acid rain. It originates from the combustion of sulfur-containing fuels, metal smelting, and petroleum refining, which produce sulfuric acid and silicate products. Given the rapid development of its economy, China has seen its highest concentrations of sulfur dioxide emissions in recent years. Three years ago, China officially promulgated several policies to measure and control sulfur dioxide pollution. As such, we believe that the prediction of sulfur dioxide emissions is important to effectively measure and control such emissions. Although many prediction techniques can be used, these approaches are relatively inaccurate because of changing environmental indicators and nonuniform standards. Thus, small simple prediction approaches are more suitable than other prediction models. In this section, three grey models are applied to predict sulfur dioxide emissions based on the data from 2007 to 2013. Table 3 lists the data from the *China Statistical Yearbook (2013)*.

Table 3 shows that Chinese sulfur dioxide emissions from 2007 to 2013 exhibit an approximately decreasing trend, which is consistent with the modeling prerequisite of the grey prediction models. The GM(1, 1) and MGM(1, 1) can be used to predict and analyze this scenario. The singular datum in 2011 is also found in this sequence. Thus, the GM(1, 1) and MGM(1, 1) could produce larger errors based on the aforementioned analysis. The IMGGM(1, 1) is then applied to generate fitting and prediction results based on this sequence. Lastly, a detailed comparison of the three grey prediction models is provided.

4.2. Prediction Analysis of the Sulfur Dioxide Emissions in China

4.2.1. Fitting and Prediction Based on the GM(1, 1). The first five numbers are considered the original data, that is, 2007- to-2011 data, and the data from 2012 to 2013 are considered the unknown data. This demarcation is the reason why the real fitting and prediction errors can be calculated and obtained and why the more suitable models can be selected for further predictions. Based on the above set,

we can obtain the original data sequence; namely, $X^{(0)} = (2468.1, 2321.2, 2214.4, 2185.1, 2217.9)$. The following calculation steps and results are first obtained by using the GM(1, 1).

Step 1. Use the 1-AGO operator, and then we have

$$X^{(1)} = (2468.1, 4789.3, 7003.7, 9188.8, 11406.7). \quad (19)$$

Step 2. Calculate the background values, and then we can obtain

$$Z^{(1)} = (3628.7, 5896.5, 8096.25, 10297.75). \quad (20)$$

Step 3. Estimate the developing coefficient $-a$ and the grey action quantity b through the least squares method, and then we have

$$B = \begin{bmatrix} -3628.7, & -5896.5, & -8096.25, & -10297.75 \\ 1, & 1, & 1, & 1 \end{bmatrix}^T, \quad (21)$$

$$Y = [2321.2, 2214.4, 2185.1, 2217.9]^T,$$

$$\alpha = (a, b)^T = (B^T B)^{-1} B^T Y$$

$$= [0.01537, 2341.927542].$$

Step 4. Calculate the time response sequence $\bar{x}^{(1)}(k+1)$, and then

$$\bar{x}_1^{(1)}(k+1) = -149904.759581e^{-0.01537k} + 152372.859581. \quad (22)$$

Step 5. Transform $\bar{x}^{(1)}(k+1)$ into the restored time response sequence $\bar{x}^{(0)}(k+1)$, and then we can obtain

$$\bar{x}^{(0)}(k+1) = \bar{x}^{(1)}(k+1) - \bar{x}^{(1)}(k)$$

$$= -149904.759581e^{-0.01537k} (1 + 149904.759581e^{0.01537}). \quad (23)$$

Based on the time response sequence $\bar{x}^{(1)}(k+1)$ and the restored time response sequence $\bar{x}^{(0)}(k+1)$, the specific calculation results are obtained using the GM(1, 1), which show in Table 4.

4.2.2. Fitting and Prediction Based on the MGM(1, 1). Based on the above results and modeling steps of the MGM(1, 1), the MGM(1, 1) is applied to fit the data of the sulfur dioxide emissions in China from 2007 to 2011 and then to predict the data from 2012 to 2013. Two metabolism calculation processes are used, the special results showed in Table 5. Here, we ignore the calculation steps and process.

TABLE 4: Fitting and prediction results based on the GM(1, 1).

Year	Actual data	Modeling data	Fitting and prediction values	Residual errors	Relative error (%)
2007	2468.1	2468.1	2468.1	0	0
2008	2321.2	2321.2	2286.3	-34.82	-1.500
2009	2214.4	2214.4	2251.5	37.10	1.6756
2010	2185.1	2185.1	2217.1	32.06	1.4674
2011	2217.9	2217.9	2183.3	-34.55	-1.5578
2012	2117.6	—	2150.0	32.44	1.5322
2013	2044.0	—	2117.2	73.25	3.5838

TABLE 5: Fitting and prediction results based on the MGM(1, 1).

Year	Actual data	Modeling data	Fitting and prediction value	Residuals	Relative errors (%)
2007	2468.1	2468.1	2468.1	0	0
2008	2321.2	2321.2	2286.3	-34.82	-1.5001
2009	2214.4	2214.4	2251.5	37.10	1.6756
2010	2185.1	2185.1	2217.1	32.06	1.4674
2011	2217.9	2217.9	2183.3	-34.55	-1.5578
2012	2117.6	—	2150.0	32.44	1.5322
2013	2044.0	—	2152.2	108.21	5.2945

4.2.3. *Fitting and Prediction Based on the IMGGM(1, 1).* Table 3 and the original data sequence show that the data from 2007 to 2010 follow a decreasing trend, whereas the 2010-to-2011 data follow an increasing trend. Therefore, the 2011 data is a singular datum. According to the aforementioned analysis, the GM(1, 1) and MGM(1, 1) produce larger errors than the IMGGM(1, 1). Therefore, the IMGGM(1, 1) is applied for further calculations and predictions in this subsection.

Step 1. Construct and apply the GM(1, 1), and obtain the time response sequence $\bar{x}_1^{(1)}(k + 1)$ and the restored time response sequence $\bar{x}_1^{(0)}(k + 1)$ as follows:

$$\begin{aligned} \bar{x}_1^{(1)}(k + 1) &= -149904.759581e^{-0.01537k} \\ &\quad + 152372.859581, \\ \bar{x}_1^{(0)}(k + 1) &= \bar{x}_1^{(1)}(k + 1) - \bar{x}_1^{(1)}(k) \\ &= -149904.759581e^{-0.01537k} (1 \\ &\quad + 149904.759581e^{0.01537}). \end{aligned} \tag{24}$$

Thus, we have $\bar{x}_1^{(0)}(2011^*) = 2183.3$, which can be taken as the fitting value of the 2011 data.

Step 2. Eliminate the singular datum $x^{(0)}(2011^*) = 2217.9$ and replace it by its first fitting data $\bar{x}_1^{(0)}(2011^*) = 2183.3$. Then, the following new original data sequence is obtained:

$$X^{(0)} = (2468.1, 2321.2, 2214.4, 2185.1, 2183.3). \tag{25}$$

The 1-AGO values of this sequence can be presented as follows:

$$X^{(1)} = (2468.1, 4789.3, 7003.7, 9188.8, 11372.1). \tag{26}$$

Here, the new original data sequence exhibits a decreasing trend, which is different from the original data sequence. Thus, this sequence is more suitable to be used in the grey prediction models.

Step 3. Construct and apply the GM(1, 1) again based on this new original sequence. The second-time time response sequence $\bar{x}_2^{(1)}(k + 1)$ and the second-time restored time response sequence $\bar{x}_2^{(0)}(k + 1)$ are obtained as follows:

$$\begin{aligned} \bar{x}_2^{(1)}(k + 1) &= -115315.891e^{-0.020089k} + 117783.991038, \\ \bar{x}_2^{(0)}(k + 1) &= \bar{x}_2^{(1)}(k + 1) - \bar{x}_2^{(1)}(k) \\ &= -115315.891e^{-0.020089k} (1 + 115315.891e^{0.020089}). \end{aligned} \tag{27}$$

Step 4. Perform the metabolism process by eliminating the fitting data $\bar{x}^{(0)}(2007)$ and metabolizing the prediction data $\bar{x}^{(0)}(2012)$, and then construct a metabolic data sequence $X^{(0)} = (x^{(0)}(2008), x^{(0)}(2009), x^{(0)}(2010), \bar{x}_1^{(0)}(2011^*), \bar{x}_2^{(0)}(2012))$, based on which construct GM(1, 1) to predict the new data in 2013. The calculation results are shown in Table 6.

Step 5. Compare the fitting and prediction errors, and then check the results which are shown and analyzed in Section 4.2.4.

TABLE 6: Fitting and prediction results based on the IMGGM(1, 1).

Year	Actual data	Modeling data	Fitting and prediction values	Residual errors	Relative errors (%)
2007	2468.1	2468.1	2468.1	0	0
2008	2321.2	2321.2	2293.4	-27.77	-1.1966
2009	2214.4	2214.4	2247.8	33.42	1.5095
2010	2185.1	2185.1	2203.1	18.03	0.8254
2011	2217.9	2183.348	2159.3	-24.01	-1.0998
2012	2117.6	—	2116.4	-1.19	-0.0565
2013	2044.0	—	2074.3	30.32	1.4835

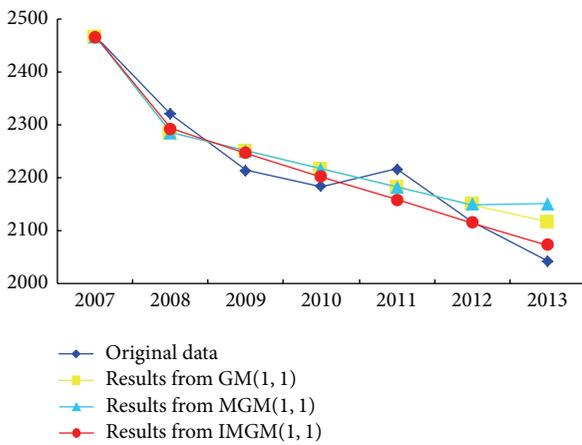


FIGURE 3: Original data and the fitting and prediction results.

4.2.4. Comparisons of the Three Grey Prediction Models. The original data and the fitting and prediction results previously provided through three grey prediction models are displayed in Figure 3 together.

Figure 3 exhibits the following conclusions:

- (i) If a singular datum exists in the original data sequence, the GM(1, 1) and MGM(1, 1) cannot effectively fit and predict the data.
- (ii) If the singular datum does not suitably fit, other fitting data after the singular datum will be inaccurate, their corresponding errors could be amplified, and the prediction results are deflected.
- (iii) The IMGGM(1, 1) can improve the fitting and prediction capability by handling singular data, which would allow for its classification as an improved grey prediction model.

The residuals and relative errors of the three grey prediction models are compared, which are shown in Figures 4 and 5.

If the residuals and relative errors are near zero, then the prediction model has high forecast accuracy. Based on this

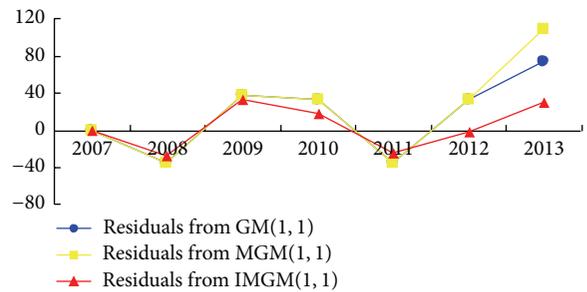


FIGURE 4: Comparisons of the residuals based on three grey prediction models.

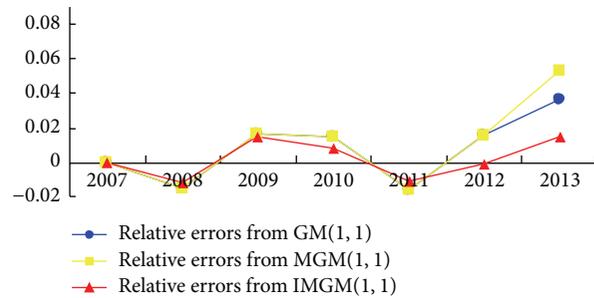


FIGURE 5: Comparisons of the relative errors based on three grey prediction models.

principle and Figures 4 and 5, the following conclusions are derived:

- (i) The singular datum exerts an inappropriate influence on the GM(1, 1) and MGM(1, 1) and caused large residuals and relative errors in the prediction results.
- (ii) The IMGGM(1, 1) can effectively transform the singular datum to a trend datum, thereby meeting the modeling prerequisite of the grey prediction model.
- (iii) The IMGGM(1, 1) has good fitting and prediction results because it has less residuals and relative errors in the prediction process than the GM(1, 1) and MGM(1, 1).

TABLE 7: Comparisons of four types of errors based on three grey prediction models.

Models	Fitting average residuals	Prediction average residuals	Fitting average relative errors	Prediction average relative errors
GM(1, 1)	27.7088	52.8512	0.00017	0.0255
MGM(1, 1)	27.7088	70.3336	0.00017	0.0341
IMGM(1, 1)	20.6504	15.761	0.00007	0.0071

TABLE 8: Prediction results of the sulfur dioxide emissions in China from 2014 to 2018.

Year	2014	2015	2016	2017	2018
Predictive value (unit: 10^4 tons)	2013.3621	1969.4262	1926.4491	1884.4098	1843.2879

By distinguishing the fitting and prediction results, we divide the residuals and relative errors into four parts: fitting average residuals, prediction average residuals, fitting average relative errors, and prediction relative errors; their calculated results for this case study are shown in Table 7.

Table 7 demonstrates the improvements and optimization of the IMGGM(1, 1) and shows that the IMGGM(1, 1) has less fitting average residuals, prediction average residuals, fitting average relative errors, and prediction relative errors than other models. Therefore, the IMGGM(1, 1) is a better grey prediction model than the GM(1, 1) and MGM(1, 1) in fitting and predicting small simple sequences with a singular datum. As such, it would be reasonable to select the IMGGM(1, 1) to predict future data on sulfur dioxide emissions in China.

4.3. *Further Prediction Based on the IMGGM(1, 1)*. Based on the results from Section 4.2.2, the prediction results of the sulfur dioxide emissions in China from 2014 to 2018 are shown in Table 8.

Table 8 shows five prediction results for sulfur dioxide emissions in China. These prediction results show *no major changes in the sulfur dioxide emission in China over the next five years* if the background emissions management and Chinese economic growth remain the same. These results also predicted that the amount of sulfur dioxide emissions in China would be 1969.4262×10^4 tons in 2015 and will be 1843.2879×10^4 tons in 2018, which in turn means that *the sulfur dioxide emissions goal proposed in the Twelfth Five-Year Plan by the China State Council will be very difficult to achieve*.

Recently, the *China Statistical Yearbook (2014)* was published by the National Bureau of Statistics of China. Thus, we have also incorporated and analyzed data on the sulfur dioxide emissions of China in 2014 and analyze it as follows.

Actual data on 2014 sulfur dioxide emission in China was 1974.4×10^4 tons; the data can be found in the *China Statistical Yearbook (2014)* or at <http://data.stats.gov.cn/easyquery.htm?cn=C01>. The prediction result based on the proposed IMGGM(1, 1) is 2013.4×10^4 tons presented (see Table 8 in this paper). Therefore, the residual error is 39×10^4 tons and the relative error is just 1.98%. This is a good result that illustrates the effectiveness of the proposed model.

5. Conclusions

This study proposed an improved grey prediction model, the IMGGM(1, 1), by combining the fitting advantage of the GM(1, 1) in small sample sequences and the additional advantages of the MGM(1, 1) in new real-time data. Given this combination, the IMGGM(1, 1) is able to overcome the limitations of both previous models when the predicted results are vulnerable at any singular datum. The IMGGM(1, 1) is suitable in modeling a small simple sequence with a singular datum and in predicting recent data. It can be an effective tool to predict economic and social data when the data are small and include a singular datum.

A practical case study on the sulfur dioxide emissions in China from 2007 to 2013, as well as the development trend, was used to illustrate the improvements and advantages of the proposed model. The results from applying the three grey prediction models lead to the following conclusions.

(1) The singular data can seriously affect the prediction effect of the GM(1, 1) and MGM(1, 1) for the modeling prerequisite.

(2) The IMGGM(1, 1) can transform a singular datum into a trend datum and thereby meet the modeling prerequisite of grey prediction models. It can also obtain better fitting and prediction results than other grey prediction models.

(3) The level of sulfur dioxide emissions in China will not decrease to the requisite level proposed in the Twelfth Five-Year Plan by the China State Council in 2015 if effective action is not taken immediately.

The IMGGM(1, 1) considers only a small simple sequence with a singular datum and does not include more singular data, which is its limitation. Although likely to be a challenging undertaking, further studies should be conducted to address this limitation.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the Natural Science Foundation of China (nos. 71301141 and 71561026), Humanity

and Social Science Youth Foundation of Ministry of Education of China (no. 13YJC630247), Science Foundation and Major Project of Educational Committee of Yunnan Province (no. 2014Z100), Applied Basic Research Programs of Science and Technology Commission of Yunnan Province (no. 2013FD029), Social Science Fund of Yunnan Province (no. YB2015087), and China Postdoctoral Science Foundation (no. 2015M570792).

References

- [1] Y.-F. Wang, "Predicting stock price using fuzzy grey prediction system," *Expert Systems with Applications*, vol. 22, no. 1, pp. 33–38, 2002.
- [2] X. Sun, W.-H. Qiu, and B.-J. Tang, "The research about model of early-warning of enterprise crisis based on gray forecasting and pattern recognition," *Systems Engineering Theory & Practice*, vol. 25, no. 5, pp. 36–42, 2005.
- [3] B. Vishnu and P. Syamala, "Grey model for stream flow prediction," *Aceh International Journal of Science and Technology*, vol. 1, no. 1, pp. 14–19, 2012.
- [4] G. W. Chang and H. J. Lu, "Forecasting flicker severity by grey predictor," *IEEE Transactions on Power Delivery*, vol. 27, no. 4, pp. 2428–2430, 2012.
- [5] D.-C. Li, C.-J. Chang, C.-C. Chen, and W.-C. Chen, "A grey-based fitting coefficient to build a hybrid forecasting model for small data sets," *Applied Mathematical Modelling*, vol. 36, no. 10, pp. 5101–5108, 2012.
- [6] H.-W. V. Tang and M.-S. Yin, "Forecasting performance of grey prediction for education expenditure and school enrollment," *Economics of Education Review*, vol. 31, no. 4, pp. 452–462, 2012.
- [7] M.-S. Yin and H.-W. V. Tang, "On the fit and forecasting performance of grey prediction models for China's labor formation," *Mathematical & Computer Modelling*, vol. 57, no. 3-4, pp. 357–365, 2013.
- [8] X. Li, X. Liao, X. Zhang, X. Chen, H. Wang, and X. Tan, "Application of GM model and GRA on the evaluation for financial burden of patients at hospitals in China by PPP model," *Journal of Grey System*, vol. 25, no. 2, pp. 112–128, 2013.
- [9] L. F. Wu, S. F. Liu, L. G. Yao, and S. L. Yan, "The effect of sample size on the grey system model," *Applied Mathematical Modelling*, vol. 37, no. 9, pp. 6577–6583, 2013.
- [10] M. Tabaszewskin and C. Cempel, "Using a set of GM(1,1) models to predict values of diagnostic symptoms," *Mechanical Systems and Signal Processing*, vol. 52–53, pp. 416–425, 2014.
- [11] H.-T. Pao, H.-C. Fu, and C.-L. Tseng, "Forecasting of CO₂ emissions, energy consumption and economic growth in China using an improved grey model," *Energy*, vol. 40, no. 1, pp. 400–409, 2012.
- [12] Z. Huang, Q. Dai, S. Lan, and S. Tie, "Optimal strategies for carbon reduction at dual levels in China based on a hybrid nonlinear grey-prediction and quota-allocation model," *Physical Review A*, vol. 83, pp. 185–193, 2014.
- [13] H. Wang, A. Luo, R. Chai, and Q. Liu, "Application of GM model in coal mine water inflow prediction," in *Proceedings of the 7th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA '15)*, pp. 192–195, Nanchang, China, June 2015.
- [14] S. Q. Ma and J. Fu, "Research of accidents due to human factor in coal mine based on GM model," *Advanced Materials Research*, vol. 1092-1093, pp. 692–695, 2015.
- [15] N.-M. Xie and S.-F. Liu, "Discrete GM(1,1) and mechanism of grey forecasting model," *System Engineering Theory and Practice*, vol. 25, no. 1, pp. 93–99, 2005.
- [16] N.-M. Xie and S.-F. Liu, "Discrete grey forecasting model and its optimization," *Applied Mathematical Modelling*, vol. 33, no. 2, pp. 1173–1186, 2009.
- [17] T. X. Yao and S. F. Liu, "Improvement of a forecasting discrete GM(1,1) model," *Systems Engineering*, vol. 25, no. 9, pp. 103–106, 2007.
- [18] Z. X. Wang, Y. G. Dang, and S. F. Liu, "An optimal GM(1,1) based on the discrete function with exponential law," *Systems Engineering—Theory & Practice*, vol. 28, no. 2, pp. 61–67, 2008 (Chinese).
- [19] C.-S. Shih, Y.-T. Hsu, J. Yeh, and P.-C. Lee, "Grey number prediction using the grey modification model with progression technique," *Applied Mathematical Modelling*, vol. 35, no. 3, pp. 1314–1321, 2011.
- [20] D.-C. Li, C.-J. Chang, W.-C. Chen, and C.-C. Chen, "An extended grey forecasting model for omnidirectional forecasting considering data gap difference," *Applied Mathematical Modelling*, vol. 35, no. 10, pp. 5051–5058, 2011.
- [21] D.-C. Li, C.-J. Chang, C.-C. Chen, and W.-C. Chen, "Forecasting short-term electricity consumption using the adaptive grey-based approach—an Asian case," *Omega*, vol. 40, no. 6, pp. 767–773, 2012.
- [22] D. Q. Truong and K. K. Ahn, "An accurate signal estimator using a novel smart adaptive grey model SAGM(1,1)," *Expert Systems with Applications*, vol. 39, no. 9, pp. 7611–7620, 2012.
- [23] J. Cui, S.-F. Liu, B. Zeng, and N.-M. Xie, "A novel grey forecasting model and its optimization," *Applied Mathematical Modelling*, vol. 37, no. 6, pp. 4399–4406, 2013.
- [24] T. D. Dai and X. M. Huang, "Selection of discrete GM model initial value by designing calculation program," *Journal of Grey System*, vol. 15, pp. 133–137, 2012.
- [25] W. Zhou and J.-M. He, "Generalized GM (1, 1) model and its application in forecasting of fuel production," *Applied Mathematical Modelling*, vol. 37, no. 9, pp. 6234–6243, 2013.
- [26] J. Lina and R.-J. Lian, "Design of a grey-prediction self-organizing fuzzy controller for active suspension systems," *Applied Soft Computing*, vol. 13, no. 10, pp. 4162–4173, 2013.
- [27] P. P. Xiong, Y. G. Dang, and T. X. Yao, "Modeling method of non-equidistant GM(1, 1) model based on optimization initial condition," *Control Decision*, vol. 11, pp. 102–109, 2015.
- [28] Z. Y. Wang, L. Chen, and K. Y. He, "PCA's application study on force made of the influence factor in coagulation," *Mathematics in Practice and Theory*, vol. 37, no. 22, pp. 60–65, 2007.
- [29] Y.-S. Lee and L.-I. Tong, "Forecasting energy consumption using a grey model improved by incorporating genetic programming," *Energy Conversion and Management*, vol. 52, no. 1, pp. 147–152, 2011.
- [30] C.-S. Lin, F.-M. Liou, and C.-P. Huang, "Grey forecasting model for CO₂ emissions: a Taiwan study," *Applied Energy*, vol. 88, no. 11, pp. 3816–3820, 2011.
- [31] L. Pan, B. Sun, and W. Wang, "City air quality forecasting and impact factors analysis based on grey model," *Procedia Engineering*, vol. 12, pp. 74–79, 2011.

- [32] W. H. Dai, X. X. Chen, and X. T. Sun, "The gray prediction model of the atmospheric pollutants in Shenzhen based on GM(1,1)," *Mathematics in Practice and Theory*, vol. 44, no. 1, pp. 131–136, 2014.
- [33] J. L. Deng, *Grey Theory*, Huazhong University of Science and Technology Press, Wuhan, China, 2002.
- [34] S. F. Liu, T. B. Guo, and Y. G. Dang, *Grey System Theory and Its Application*, Science Press, Beijing, China, 3rd edition, 2004.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

