Professor Che-Jung CHANG, PhD
E-mail: r3795102@nckualumni.org.tw
TSL Business School, Quanzhou Normal University
Fujian University Engineering Research Center of Cloud Computing Internet of Things and E-Commerce Intelligence
Associate Professor Guiping LI, PhD
E-mail: liguipingde@126.com
Department of Management Science and Engineering, Business School Ningbo University
Associate Professor Jianhong GUO*, PhD
E-mail: jianhongguo@126.com
Associate Professor Kun-Peng YU, PhD
E-mail: levinroo@126.com
TSL Business School, Quanzhou Normal University
Fujian University Engineering Research Center of Cloud Computing Internet of Things and E-Commerce Intelligence

DATA-DRIVEN FORECASTING MODEL FOR SMALL DATA SETS

Abstract. *The effective use of information is an important foundation for a company's sustainable development and effective management of its operations. However, when making important management decisions, managers often face issues of insufficient information or limited data due to time or cost constraints. The use of a grey model is a common solution for solving this small-data-set issue; this model has been successfully applied to various fields with reliable outcomes. Nevertheless, it may not always achieve sufficient accuracy, especially in predicting non-equigap data. This study introduces the concept of fuzzy membership functions to reform the formula of the background values of the data that is input, and then proposes a data-driven grey model for a small amount of non-equigap data. Two real case studies involving material-fatigue-limit testing data and the monthly demand for a specific uninterruptible power supply product are taken as examples to demonstrate the proposed method. The experimental results show the proposed method is able to obtain a solid outcome, yielding accurate forecasts using a small amount of non-equigap data.*

Keywords*: Forecasting, Small data set, Grey theory, Non-equigap, Short-term demand, Fatigue limit.*

JEL Classification: C01, C02, C13, C22, C53, M11

* Corresponding author.

DOI: 10.24818/18423264/54.4.20.14

1. Introduction

In the contemporary knowledge economy, the effective use of information has become essential for the sustainable development of businesses (Liu & Lin, 2006). However, because of time or cost constraints, the data available to managers is often limited. Managers frequently must make important decisions using uncertain information. Therefore, a forecasting method based on limited data is need to help enterprises enhance their competitiveness.

To solve the uncertainty problem, various theories have been suggested, such as probability statistics (Lind et al., 2008), fuzzy mathematics (Kilr & Yuan, 1995), and the grey system theory (Deng, 1982). In theory, probability statistics may be used to solve random phenomena; fuzzy mathematics may be used to solve cognitive uncertainty problems; and the grey system theory focuses on small-data-set analysis problems. In practice, the use of probability statistics requires a large number of samples to produce better results, while fuzzy mathematics relies on human experience. These characteristics make these two methods unsuitable for small-data-set analysis. The grey system theory, however, is a viable option for solving the problem of insufficient information.

With the help of the grey system theory, many otherwise unsolvable small-data-set analysis problems can be resolved successfully (Xie & Liu, 2009). Grey prediction is a primary branch of grey system theory. The first-order one-variable grey model, GM(1,1), is the main model of grey prediction; it can be built with a small number of observations (four or more) and still produce a reliable forecasting result (Liu & Lin, 2006; Wu et al., 2013). It depends on an accumulating generation operator (AGO) to find the inherent regular pattern from a data set, and it is one of the most widely used techniques in terms of grey system theory. In recent decades, it has been successfully utilized in various fields with satisfactory results (Boran, 2015; Salehi & Dehnavi, 2018; Talafuse & Pohl, 2018; Yamaguchi et al., 2007; Zeng et al., 2018).

However, the GM(1,1) has been developed for time series data with equigap features; this limits its use. To extend the applicability of the model, many adjusted models have been proposed that do allow the grey model to handle non-equigap data (Dai & Li, 2005; He & Sun, 2001; Shi, 1993), but the predictive performance of these new models can be further improved.

Other than the grey system theory models, the virtual sample generation (VSG) technique is another mainstream method for small-data-set forecasting expect (Li et al., 2003). VSG generates artificial samples based on prior knowledge acquired from the original small training set, and it is a useful supplement for improving the learning performance (Li & Lin, 2006). However, this technique is not directly applicable to time series data; developing this data type is closely related to the order of observations, so it is difficult to maintain the fitting relations among the generated artificial data, which implies that the VSG technique does not effectively improve the learning performance (Chang et al., 2015).

DOI: 10.24818/18423264/54.4.20.14

While there are limitations on the use of VSG, it still provides some advanced ideas for solving small-data-set forecasting problems. The general principle of small-data-set handling is to fully develop and utilize the existing data (Lin & Tsai, 2014). If the derived information from the existing small data set is utilized properly, it is possible to enhance the processing performance for small samples (Chang et al., 2019).

Based on this concept, this study adopts one of the VSG techniques (Li et al., 2007)— the mega-trend-diffusion (MTD) procedure— to analyze data features and extract latent information to reform the non-equigap grey model. The approach first uses the MTD procedure to estimate the possible range of data and then establishes a membership function (MF) based on the estimated range. The corresponding MF values that are obtained yield important location information for determining the alpha coefficient, which is a key parameter for computing the background values that are used in the grey modeling process to improve forecasting accuracy. The proposed method based on the MTD procedure, which we have termed MTD-NGM(1,1), is able to establish a model according to the data features and is a feasible data-driven forecasting model for non-equigap data.

This study utilizes two real case studies to demonstrate the validity of the proposed method in an empirical comparison. The first case is based on fatigue-limit test data selected from related literature (Chang et al., 2015), and it discusses the relationship between temperature changes and fatigue limits. The second case is provided by a global leading manufacturer in switching power supply. The figures for monthly demand for a specific uninterruptible power supply (UPS) product are applied to construct forecasting models, and the performance of the models in solving short-term demand predicting problems is scrutinized. The empirical results show that the proposed method improves forecasting accuracy and thus is suitable for small-data-set problems.

The remainder of this paper is organized as follows: Section 2 introduces the concept and calculation steps of the proposed method are introduced; Section 3 describes and examines the performance of the proposed method; Section 4 presents the paper's conclusions.

2. Methodology

The primary challenge is extracting enough useful information for modeling when the obtained sample size is small. Taking this into account, this work utilizes the MTD technique to extend the domain range and then adopts the MF to find the location information of each piece of data. Next, we develop a modified modeling procedure, called MTD-NGM(1,1), to increase the forecasting performance of the grey approach. This section will explain the concept of the MTD-NGM(1,1) and its executable process.

DOI: 10.24818/18423264/54.4.20.14

2.1 Domain range estimation

The MTD method was proposed by Li et al. (2007), and it is a technique for estimating the domain range of small samples. It fills the data gap based on possible trends in the data. Its basic assumption is that samples will fall within a certain range and have their own location information. MTD involves a heuristic mechanism that fills possible data gaps. This method adopts the central tendency, dispersion, and skewness characteristics to describe data features.

In the MTD procedure, Equations (1) and (2) are used to estimate the upper bound (UB) and lower bound (LB) of a data set, where $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of the acquired observations, and $CL = \left(x_{\max} + x_{\min}\right)/2$ is the central location. $\hat{s}_x^2 = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 / \left(n-1\right)$ is the variance of observations; n is the sample size; $\left|X^+\right|$ and $\left|X^-\right|$ represent the number of observations that are greater and smaller than the CL, respectively. Equations (1) and (2) are the original settings for the upper and lower bounds, and $\ln\left(10^{-20}\right)$ is called the diffusion coefficient. However, to avoid the phenomenon of insufficient expansion under special circumstances, the formula is adjusted to Equations (3) and (4).

$$UB = CL + \frac{\left|X^+\right|}{\left|X^+\right| + \left|X^-\right|} \times \sqrt{-2 \times \hat{s}_x^2 / \left|X^+\right| \times \ln\left(10^{-20}\right)} \quad (1)$$

$$LB = CL - \frac{\left|X^-\right|}{\left|X^+\right| + \left|X^-\right|} \times \sqrt{-2 \times \hat{s}_x^2 / \left|X^-\right| \times \ln\left(10^{-20}\right)} \quad (2)$$

$$UB_{adj} = \begin{cases} UB = CL + \frac{\left|X^+\right|}{\left|X^+\right| + \left|X^-\right|} \times \sqrt{-2 \times \hat{s}_x^2 / \left|X^+\right| \times \ln\left(10^{-20}\right)} & UB \geq x_{\max} \\ x_{\max} & UB < x_{\max} \end{cases} \quad (3)$$

$$LB_{adj} = \begin{cases} LB = CL - \frac{\left|X^-\right|}{\left|X^+\right| + \left|X^-\right|} \times \sqrt{-2 \times \hat{s}_x^2 / \left|X^-\right| \times \ln\left(10^{-20}\right)} & LB \leq x_{\min} \\ x_{\min} & LB < x_{\min} \end{cases} \quad (4)$$

2.2 Location information

The MTD procedure adopts the concept of fuzzy theory to reconstruct the distribution of the observations. As shown in Figure 1, UB_{adj}, LB_{adj}, and CL are used to form an asymmetric triangular MF. The corresponding MF values are called location information, which represent possibility indexes of sample occurrences. The location information presents the degree of closeness between a single datum and the central location (CL), and it can reflect the importance level

DOI: 10.24818/18423264/54.4.20.14

of a single datum. The MTD procedure estimates the possible data profile and obtains useful location information for sparse data. This study introduces this location information, the MF values, to develop an improved grey modeling procedure. The following are the formulas of the MF.

$$MF = \begin{cases} (x_i - LB)/(CL - LB) & x_i \le CL \\ (UB - x_i)/(UB - CL) & x_i > CL \end{cases} \tag{5}$$

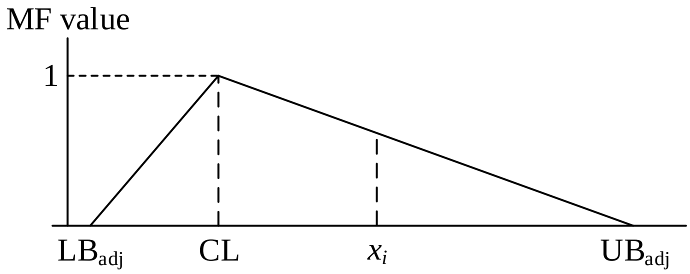

Figure 1. Asymmetric triangular MF

2.3 Model basis: NGM(1,1)

The conventional grey model generates a smoother series to build a model through the AGO based on equigap data. However, the practical data is not always equigap, so its application scope is impeded by the building principle. To overcome this dilemma, He and Sun (2001) developed a new formula in the AGO to allow the grey model to applied to non-equigap data. This model, NGM(1,1), maintains easy-to-use features and is also one of the grey non-equigap models with high accuracy. Therefore, this study proposes an improved modeling approach based on the NGM(1,1).

The modeling process of the NGM(1,1) is described as follows:

Step 1: Given the original time series data set $X^{(0)} = \{x^{(0)}(k_1), x^{(0)}(k_2), \ldots, x^{(0)}(k_n)\}$, $n \ge 4$, and $x^{(0)}(k_i)$ represents the corresponding datum index of the *i*th datum. Let $\Delta k_i = k_i - k_{i-1}$, and $i = 2,3,\ldots,n$ is the gap of each paired datum index.

Step 2: Transform the original series $X^{(0)}$ into a new series $X^{(1)} = \{x^{(1)}(k_1), x^{(1)}(k_2), \ldots, x^{(1)}(k_n)\}$ by the AGO.

$$x^{(1)}(k_1) = x^{(0)}(k_1);\ x^{(1)}(k_i) = x^{(0)}(k_1) + \sum_{j=2}^{i} x^{(0)}(k_j)\Delta k_j,\ i = 2,3,\ldots,n \tag{6}$$

Step 3: Compute the background values $z^{(1)}(k)$ using Equation (7).

$$z^{(1)}(k) = (1-\alpha)x^{(1)}(k_{i-1}) + \alpha x^{(1)}(k_i),\ \alpha \in (0,1),\ k = 2,3,\ldots,n \tag{7}$$

DOI: 10.24818/18423264/54.4.20.14

Step 4: Estimate the developing coefficient a and the grey input b by the least-squares method from Equation (8). The estimated vector, $\hat{\mathbf{a}} = [a,b]^T$, can be computed by Equation (9).

$$x^{(0)}(k) + az^{(1)}(k) = b \tag{8}$$

$$[a,b]^T = \left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{Y} \tag{9}$$

where

$$\mathbf{B} = \begin{bmatrix} -z^{(1)}(k_2) & 1 \\ -z^{(1)}(k_3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(k_n) & 1 \end{bmatrix} \tag{10}$$

$$\mathbf{Y} = \left[x^{(0)}(k_2), x^{(0)}(k_3), \ldots, x^{(0)}(k_n)\right]^T \tag{11}$$

Step 5: Establish the first order grey differential equation by Equation (12).

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \tag{12}$$

Step 5: Use the estimated vector together with the initial condition $x^{(0)}(k_1) = x^{(1)}(k_1)$ to solve Equation (12), and then the desired forecasting output can be extracted by Equation (13).

$$\begin{cases} \hat{x}^{(1)}(k_i) = \left(x^{(0)}(k_1) - \dfrac{b}{a}\right)e^{-a(k_i - k_1)} + \dfrac{b}{a} \\ \hat{x}^{(0)}(k_i) = \dfrac{\hat{x}^{(1)}(k_i) - \hat{x}^{(1)}(k_{i-1})}{\Delta k_i} \end{cases} \tag{13}$$

2.4 Model basis: NGM(1,1)

This subsection describes the proposed method and its modeling procedure.

2.4.1 Influence of the alpha coefficient in grey modeling

The background value is an important aspect of grey modeling, as it directly affects the estimated coefficients and impacts the forecasting performance. The background value serves two functions: the first is smoothing the data to alleviate randomness; the second is emphasizing the importance of the newest datum. The alpha coefficient plays a key role in the calculation of the background value (Li et al., 2009).

The alpha coefficient is usually set to a fixed value (0.5) for easy use, which regards all data as equivalence to simplify the complex calculations.

DOI: 10.24818/18423264/54.4.20.14

However, this approach may overlook useful information that results from differences in the data features, and it may lack adaptability for different data types, thereby causing significant errors. When Equation (7) is rewritten as Equation (14), it is clear that the alpha coefficient acts only on the newest datum when the background value is calculated (Li et al., 2009). The alpha value thus may be considered to represent the importance of the newest datum. Hence, it is essential to develop a data-driven systematic approach to determine the proper alpha coefficient.

$$z^{(1)}(k)=x^{(1)}(k-1)+\alpha x^{(0)}(k),\ \alpha\in(0,1),\ k=2,3,\ldots,n \tag{14}$$

2.4.2 Modelling process

The MF value of each datum is the location information obtained through data behavior analysis. Figure 1 shows that the location information could represent the relative proximity level between the individual datum and the CL, i.e. the typical degree of a single datum in the dataset. The more typical data in the dataset will have less random fluctuation, and the modeling process should emphasize the role of the newest data; otherwise, it should focus on the reduction of the random fluctuation. Its purpose likes the effect of the alpha coefficient for the background value. Therefore, the location information is spread to compute the alpha coefficient. Further, to consider the difference and importance of the data order, the data order is used as a weight, and the proposed alpha coefficient is determined from the weighted average of the MF values.

The following is the modeling process of the proposed MTD-NGM(1,1).

Step 1: Given n time series data $X^{(0)}=\left\{x^{(0)}(k_1),x^{(0)}(k_2),\ldots,x^{(0)}(k_n)\right\}$, where k_i represents the corresponding datum index of ith datum, let $\Delta k_i=k_i-k_{i-1}$, and $i=2,3,\ldots,n$ is the gap of each paired datum index.

Step 2: Compute the MF values of the existing data by the MTD procedure (Subsection 2.2) $MF=\left\{MF_1,MF_2,\ldots,MF_n\right\}$.

Step 3: Utilize Equation (6) to form a new increasing series $X^{(1)}=\left\{x^{(1)}(k_1),x^{(1)}(k_2),\ldots,x^{(1)}(k_n)\right\}$.

Step 4: Determine the alpha coefficient by Equation (15).

$$\alpha_i=\frac{\sum_{j=1}^{i} j\times MF_j}{\sum_{j=1}^{i} j},\ i\geq 2 \tag{15}$$

Step 5: Calculate the background values by Equation (16).

$$z^{(1)}(k_i)=x^{(1)}(k_{i-1})+\alpha_i\times x^{(0)}(k_i),\ i=2,3,\ldots,n \tag{16}$$

Step 6: Establish the grey differential equation and estimate the grey coefficient a and the grey input b by Equations (8)-(12).

DOI: 10.24818/18423264/54.4.20.14

Step 7: Construct the grey model by Equation (13) to obtain the desired predictions.

3. Experimental study

In this section, two real forecasting problems are described, and then the experiment is implemented; finally, a performance comparison is exhibited.

3.1 Material-fatigue-limit testing data

The material-fatigue-limit testing data is selected from a previous study to demonstrate the performance of the proposed method. The data concern the relationship between the fatigue limit of titanium alloy and temperature changes that occur under sustained symmetrical cyclic stresses.

Fatigue is a phenomenon that an external force exerts on material components until breakage occurs under a certain cyclic stress. Fatigue is the main factor that causes material components to rupture, so analyzing fatigue behavior is crucial for safety. For example, if aircraft engineers fail to consider material fatigue, the accident rates of the planes or components that they design may increase. Fatigue life and fatigue limit are two important factors to consider when analyzing fatigue behavior (Callister & Rethwisch, 2008). Understanding the material fatigue limit allows engineers to effectively utilize the available material and prevent the potential risks of fatigue fracture. However, the fatigue-limit test is usually not iterative, because it destroys the materials being tested, so the samples used for fatigue-limit tests must be small. Consequently, effectively using small samples to accurately determine the fatigue strength of materials is crucial.

Table 1 lists the obtained data. There are two variables in this data set— temperature (measured in degrees Celsius) and limiting stress level σ_{-1} (measured in mega-pascal, MPa) and they are regarded as independent and dependent variables, respectively. It would be optimal to forecast the fatigue limit based on temperature. This case study is used to confirm the effectiveness of the proposed method for making accurate predictions using a small amount of non-equigap data. Therefore, only four samples are utilized to build a model to forecast the limiting stress levels under various temperatures.

Table 1. Material-fatigue-limit testing data

^{o}C	100	130	170	210	240	270	310	340	380
σ_{-1}	560	557.54	536.1	516.1	505.6	486.1	467.4	453.8	436.4

3.2 UPS product demand

The UPS product exists in a mature industry— barrier to entry is not high and thus there is increasingly fierce competition. In recent years, the profits and

DOI: 10.24818/18423264/54.4.20.14

prices of UPS products have been declining. Effective operational management has become an important way to maintain a competitive advantage for manufacturers. An appropriate production plan can stabilize product supply and demand and is necessary for effective production management. However, uncertainty exists in the planning process generally and is a pending issue for managers when planning a project. Prediction is a feasible solution to ease the uncertainty involved in planning, as it provides information about future consumer demand, thereby helping managers to develop an appropriate project plan. If production rates can be coordinated with market demand, supply and demand can be predicted and balanced, enabling effective operational management (Krajewski et al., 2010).

Further, profits are decreasing and labor costs are increasing every year in the UPS industry. A more accurate prediction methodology will help managers schedule ongoing operational processes such as capacity arrangement and manpower schedules. Thus, forecasting short-term demand is necessary for determining the daily production plan in terms of production management. However, the prediction period is short, and the variations in demand are relatively large. It is difficult to understand the developing patterns of demand using a large number of historical observations due to the instability in the demand trends. Using recent observations with the latest information to build a model could reflect an actual situation correctly and produce accurate predictions. An accurate short-term demand forecast allows enterprises to maintain their competitive advantages and conduct efficient and effective production management.

In this study, to assess the performance of the proposed grey model, the second type of data selected is provided by a global leading manufacturer in switching power supply. The data set consists of data concerning the monthly demand for one of this company's UPS product; the dataset used comprises 12-period non-equigap time-series data from July 2011 to December 2012. In the 18 months of data used, there are six months in which no demand occurred. In addition, all collected data are converted to fall within [1, 2] by using minimum-maximum standardization for confidentiality requirements. The standardized data are listed in Table 2. In the experiment, we use four pieces of data to build a model for predicting the next observation, $\hat{x}\left(k_5\right)$.

Table 2. Monthly demand for a specific UPS product

Months	Demands	Months	Demands	Months	Demands
2011/07	1.492	2012/01	1.207	2012/07	1.000
2011/08	1.706	2012/02	2.000	2012/08	-
2011/09	-	2012/03	-	2012/09	1.037

DOI: 10.24818/18423264/54.4.20.14

2011/10	1.198	2012/04	1.465	2012/10	-
2011/11	1.412	2012/05	1.385	2012/11	1.037
2011/12	-	2012/06	-	2012/12	1.144

3.3 UPS product demand

Here the first four pieces of data in the material-fatigue-limit testing dataset are used as an example to exhibit the modeling details of the proposed method. The computation is described as follows.

1. The original data set is $X^{(0)}=\{560, 557.54, 536.1, 516.1\}$, and the data index of these data is $\{k_1, k_2, k_3, k_4\}=\{100, 130, 170, 210\}$.

2. Compute the MF values by the MTD and get $MF=\{0.6856, 0.7209, 0.9721, 0.6856\}$.

3. Utilize AGO to form a new series $X^{(1)}=\{560, 17286.2, 38730.2\ 59374.2\}$.

4. Determine the coefficient $\alpha_k=\{0.7091, 0.8406, 0.7786\}$.

5. Calculate the background values $Z^{(1)}=\{12420.868, 35311.931, 54803.819\}$.

6. Compute $\hat{\mathbf{a}}=\left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{Y}$, where $Y=\left[557.54, 536.1, 516.1\right]^T$ and $B=\begin{bmatrix} -12420.87 & 1 \\ -35311.93 & 1 \\ -54803.82 & 1 \end{bmatrix}$. Thus, $\hat{\mathbf{a}}=\left[0.000977, 569.957929\right]^T$.

7. Use Equation (11) to build the model $x^{(1)}\left(k_i\right)=\left(-582815.567\right)e^{-0.000977(k_i-100)}+583375.567$; the next corresponding output is predicted as $\hat{x}_{540}=503.9956$.

3.4 Comparisons

Accuracy is an important criterion used to measure forecasting performance (Yokum & Armstrong, 1995) and to select a forecasting strategy (Dalrymple, 1987). Liu and Lin (2006) pointed out that it is necessary to use forecasting errors to test whether a grey model is functioning properly. Therefore, we select the mean absolute percentage error (MAPE) to evaluate the performance of the proposed method. The MAPE is a relative percentage of errors, and it provides a useful tool for mangers for understanding the risks of adopting the

DOI: 10.24818/18423264/54.4.20.14

forecasting tool. y_i and $\hat{y}_i$ are the actual observation and the forecasting output, respectively, and the MAPE's formula is given as Equation (17).

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\% \quad (17)$$

The forecasting results will be compared with one typical grey non-equigap model and two popular forecast approaches; these are NGM(1,1), the support vector regression (SVR), and the back-propagation neural network (BPNN). The first is the theoretical basis of the proposed method, which was proposed by He and Sun (2001). The NGM(1,1) is one of most accurate grey non-equigap models currently in use. The SVR is a non-parametric estimation learning algorithm based on statistical learning theory, and it can be used to solve training problems with limited data. The BPNN is widely adopted in processing non-linear data. The building parameter alpha in He and Sun's NGM(1,1) is fixed at 0.5 in the experiments.

The comparison in Table 3 shows that the proposed method could produce a better forecasting outcome than the other three approaches. The improvements between the original NGM(1,1) and the proposed method are 17.33% and 9.59% in these two datasets, respectively. This demonstrates that introducing the MTD procedure could overcome the shortcomings of NGM(1,1) and achieve better results. These results reveal that the proposed MTD-NGM(1,1) transcends the other three approaches, and that the proposed method is a feasible tool for solving small non-equigap forecasting problems.

Table 3. MAPE(%) of various models

Methods	Material-fatigue-limit data	UPS demand data
Proposed method	0.9695	21.0241
NGM(1,1)	1.1727	23.2541
SVR	1.1193	22.9684
BPNN	2.1154	29.4532

4. Conclusion and discussion

In a competitive operating environment, the effective use of information is fundamental to maintaining competitiveness. However, due to time and cost constraints, the data available to managers is often limited. The grey system theory focuses on small-data-set analysis and is an option to solve this problem.

Since the grey system theory was proposed, grey prediction models have been successfully applied in many fields. However, they may not always yield accurate predictions, especially when non-equigap data is used. Therefore, this

DOI: 10.24818/18423264/54.4.20.14

study proposes a novel procedure, named MTD-NGM(1,1), to enhance the model's prediction ability for small non-equigap data. This paper applies material-fatigue-limit testing data and UPS demand data to assess the proposed model's performance. The proposed method is found to produce more accurate predictions than three other methods. These results reveal the practical application value of the proposed method.

In the future, the proposed method may be applied to other fields, such as finance, energy, and transportation to further confirm its effectiveness. Integrating an optimization algorithm into the grey modeling process is also another potential future research direction.

ACKNOWLEDGMENTS

This research was supported by Fujian Provincial Social Science Planning Project of China under Grant No. FJ2019B099, Zhejiang Provincial Natural Science Foundation of China under Grant No. LY19G010002, Qianjiang Talent Program of Zhejiang Province (China), and Humanistic and Social Science Youth Foundation of the Ministry of Education of China under Grant No. 17YJC630204.

REFERENCES

[1] **Boran, F.E. (2015),** ***Forecasting Natural Gas Consumption in Turkey Using Grey Prediction.*** *Energy Sources Part B-Economics Planning and Policy,* 10, 208-213;
[2] **Callister, W.D., Rethwisch, D.G. (2008),** ***Fundamentals of Materials Science and Engineering: An Integrated Approach;*** *John Wiley & Sons;*
[3] **Chang, C.-J., Li, D.-C., Chen, C.-C., Wen-Chih, C. (2019),** ***Extrapolation-based Grey Model for Small-data-set Forecasting.*** *Economic Computation and Economic Cybernetics Studies and Research; ASE Publishing,* 53, 171-182;
[4] **Chang, C.J., Dai, W.L., Chen, C.C. (2015),** ***A Novel Procedure for Multimodel Development Using the Grey Silhouette Coefficient for Small-Data-Set Forecasting.*** *Journal of the Operational Research Society,* 66, 1887-1894;
[5] **Dai, W.I., Li, J.F. (2005),** ***Modeling Research on Non-equidistance GM(1,1) Model.*** *Systems Engineering - Theory & Practice,* 25, 89-93;
[6] **Dalrymple, D.J. (1987),** ***Sales Forecasting Practices: Results from a United States survey.*** *International Journal of Forecasting,* 3, 379-391;
[7] **Deng, J.L. (1982),** ***Control Problems of Grey Systems.*** *Systems & Control Letters,* 1, 288-294;
[8] **He, X.G., Sun, G.Z. (2001),** ***A Non-equigap Grey Model NGM(1,1).*** *Journal of Grey System,* 13, 189-192;
[9] **Kilr, G.J., Yuan, B. (1995),** ***Fuzzy Sets and Fuzzy Logic: Theory and Applications;*** *Pearson Education;*

DOI: 10.24818/18423264/54.4.20.14

[10] **Krajewski, L.J., Pitzman, L.P., Malhotra, M.K. (2010),** ***Operations Management: Processes and Supply Chains;*** *Pearson Education, Limited;*
[11] **Li, D.C., Chen, L.S., Lin, Y.S. (2003),** ***Using Functional Virtual Population as Assistance to Learn Scheduling Knowledge in Dynamic Manufacturing Environments.*** *International Journal of Production Research,* 41, 4011-4024;
[12] **Li, D.C., Lin, Y.S. (2006),** ***Using Virtual Sample Generation to Build up Management Knowledge in the Early Manufacturing Stages.*** *European Journal of Operational Research,* 175, 413-434;
[13] **Li, D.C., Wu, C.S., Tsai, T.I., Lina, Y.S. (2007),** ***Using Mega-trend-diffusion and Artificial Samples in Small Data Set Learning for Early Flexible Manufacturing System Scheduling Knowledge.*** *Computers & Operations Research,* 34, 966-982;
[14] **Li, D.C., Yeh, C.W., Chang, C.J. (2009),** ***An Improved Grey-based approach for Early Manufacturing Data Forecasting.*** *Computers & Industrial Engineering,* 57, 1161-1167;
[15] **Lin, Y.S., Tsai, T.I. (2014),** ***Using Virtual Data effects to Stabilize Pilot Run Neural Network Modeling.*** *Journal of Grey System,* 26, 84-94;
[16] **Lind, D.A., Marchal, W.G., Wathen, S.A. (2008),** ***Basic Statistics for Business & Economics;*** *McGraw-Hill;*
[17] **Liu, S.F., Lin, Y. (2006),** ***Grey Information: Theory and Practical Applications;*** *Springer-Verlag;*
[18] **Salehi, M., Dehnavi, N. (2018),** ***Audit Report Forecast: An Application of Nonlinear Grey Bernoulli Model.*** *Grey Systems: Theory and Application,* 8, 295-311;
[19] **Shi, B.Z. (1993),** ***Modeling of Non-equigap GM(1,1).*** *Journal of Grey System,* 5, 105-113;
[20] **Talafuse, T.P., Pohl, E.A. (2018),** ***Small Sample Discrete Reliability Growth Modeling Using a Grey Systems Model.*** *Grey Systems: Theory and Application,* 8, 246-271;
[21] **Wu, L.F., Liu, S.F., Yao, L.G., Yan, S.L. (2013),** ***The Effect of Sample Size on the Grey System Model.*** *Applied Mathematical Modelling,* 37, 6577-6583;
[22] **Xie, N.M., Liu, S.F. (2009),** ***Discrete Grey Forecasting Model and its Optimization.*** *Applied Mathematical Modelling,* 33, 1173-1186.
[23] **Yamaguchi, D., Li, G.D., Nagai, M. (2007),** ***A Grey-based Rough Approximation Model for Interval Data Processing.*** *Information Sciences,* 177, 4727-4744;
[24] **Yokum, J.T., Armstrong, J.S. (1995),** ***Beyond Accuracy: Comparison of Criteria Used to Select Forecasting Methods.*** *International Journal of Forecasting,* 11, 591-597;
[25] **Zeng, B., Tan, Y., Xu, H., Quan, J., Wang, L., Zhou, X. (2018),** ***Forecasting the Electricity Consumption of Commercial Sector in Hong Kong Using a Novel Grey Dynamic Prediction Model.*** *Journal of Grey System,* 30, 159-174.

DOI: 10.24818/18423264/54.4.20.14