PREDICTING THE NUMBER OF TEACHERS AND STUDENTS AT THE ETHNIC MINORITY BOARDING SCHOOLS IN THE MEKONG DELTA REGION, VIETNAM

Thieu Van Nam [1], Nguyen Phuoc Hai [2]

[1] *Kien Giang Department of Education and Training, Vietnam*
[2] *Kien Giang Teachers Training College, Vietnam*

ABSTRACT

The purpose of this study is to predict the number of teachers and students at the ethnic minority boarding schools in the Mekong Delta region, Vietnam based on Taylor approximation method in grey model. This prediction model can obtain the most optimal predicted values by multi-times approximate calculation to improve the predicted accuracy of grey model. In addition, researchers have used the MATLAB software to build a MATLAB toolbox for this prediction model. The results of this study will provide important information for educational administrators to propose an appropriate policy for building education development strategies that are appropriate for the new conditions.

Keywords: *ethnic minority boarding, Mekong Delta, Taylor approximation method, grey prediction model, MATLAB toolbox.*

1. INTRODUCTION

The Vietnamese educational system nowadays comprises five levels: preschool education (early childhood education), primary education, secondary education (lower secondary education and upper secondary education), higher education and postgraduate education. In addition, vocational education and training provides educational opportunities for those secondary school leavers who are unable to enter higher education (Doan, 2005). Ethnic boarding school is a school model that was born before the requirements of educational practice, through the process of formation and development has proven its superiority and efficiency in the implementation of education equality, developing secondary education, improving people's knowledge, providing human resources for labor directly to locality. In particular, the boarding school for ethnic minorities also performs the mission of creating resources for ethnic minority students to receive higher education and professionalism, providing a highly qualified workforce, meeting the socio-economic development requirements of each locality where ethnic minorities live. Predicting the exact number of teachers and students at the ethnic minority boarding schools in the Mekong Delta region, Vietnam will provide important information for educational administrators to propose an appropriate policy for building education development strategies that are appropriate for the new conditions. In this paper, researchers have used Taylor approximation method in grey model to predict the number of teachers and students at the ethnic minority boarding schools in the Mekong Delta region, Vietnam.

In 1982, grey system theory was initially presented by Deng (Deng, 1989). It included five major parts that are grey prediction, grey relation, grey decision, grey programming, and grey control (Li et al., 2005). Grey model is an important part of grey system theory. Grey model has become an effective method to study uncertainty problems under discrete data and incomplete information (Deng et al., 2012). In recent years, grey model has been successfully applied to many prediction

fields as engineering, economics, medicine. The advantage of grey model is that it only needs a small amount of data and random sample data to calculate and give prediction results (Chen et al., 2012). Grey model with the differential equation is established on the basis of discrete data. It should be noticed that the minimum number of sample data in the grey model is four. However, many researchers have pointed that there were some problems occurred that the predicted accuracy of grey model was unsatisfied (Li et al., 2007; Wen and Chang, 2005), the parameters of prediction model based on grey model were not the optimal parameters, and the prediction precision of the model was not stable, they have performed a lot of researches for this to improve the predicted accuracy (Chen, 2008; Zeng et al., 2010; Pai et al., 2011; Zhou, 2013). In 2014, Nguyen and co-workers proposed using the T-GM(1,1) model and T-GM(1,*n*) model to predict the number of teachers and students for admission. Sheu and co-workers (2014) proposed using the T-GM(2,1) model to predict the number of students for admission. Nguyen et al., (2017) proposed the prediction of the admission teacher's number in Taiwan by using T-GM(1,*n*) and T-GM(2,*n*) method. Nguyen et al., (2019) proposed a new prediction model T-DGM(2,*n*) and its application to predict the number of teachers for admission in Taiwan. In this paper, researchers have used Taylor approximation method in grey model to predict the number of teachers and students at the ethnic minority boarding schools in the Mekong Delta region, Vietnam.

The remainder of this paper is organized as follows: Section 2 explains the basic theories, including grey model (GM(1,1)) and Taylor approximation method in grey model (T-GM(1,1)); Building the MATLAB toolbox for T-GM(1,1) is introduced in Section 3; Section 4 describes results and discussion for using T-GM(1,1) to predict the number of teachers and students; Finally, conclusions are presented in Section 5.

2. BASIC THEORIES

2.1 Grey Model (GM(1, 1))

Before using grey model, the initial data have to be tested based on Equation (1). The initial data have $m \geq 4$, $x^{(0)} \in R^{+}$, and

$$\left.\begin{aligned} \sigma^{(0)}(k) &\in \left(e^{-\frac{2}{m+1}}, e^{\frac{2}{m+1}} \right) \\ \sigma^{(0)}(k) &= \frac{x^{(0)}(k-1)}{x^{(0)}(k)} \end{aligned}\right\} \tag{1}$$

where $k = 2,3,\cdots,n$; $\sigma^{(0)}(k)$ is called class ratio. In grey system theory, GM(1,1) is one of the most widely used model (Nguyen et al., 2014). In general, GM(1,1) requires four observations to build a grey prediction model. Assume that $x^{(0)}$ is the original sequence as follows.

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(m)) \tag{2}$$

where *m* is the sequence length. $x^{(1)}$ is the 1-AGO sequence of $x^{(0)}$ as

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(m)) \tag{3}$$

where $x^{(1)}(1) = x^{(0)}(1)$, and

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), \; k = 1,2,\cdots,m \tag{4}$$

GM(1,1) can be constructed by establishing a first order differential equation for $x^{(1)}(k)$ as

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \tag{5}$$

Then, by least-square method, the coefficients *a* and *b* can be obtained as

$$\hat{a} = [a,b]^T = (B^T B)^{-1} B^T Y \tag{6}$$

$$\text{where} \quad B = \begin{bmatrix} -0.5(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -0.5(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -0.5(x^{(1)}(m-1) + x^{(1)}(m)) & 1 \end{bmatrix} \tag{7}$$

$$Y = \left[x^{(0)}(2), x^{(0)}(3), \cdots, x^{(0)}(m)\right]^T \tag{8}$$

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a} \tag{9}$$

The result obtained $\hat{x}^{(1)}$ from Equation (9). Applying the inverse accumulated generation operation (IAGO). The predicted equation is

$$\hat{x}^{(0)}(k+1) = (x^{(0)}(1) - \frac{b}{a})(1 - e^{a})e^{-ak} \tag{10}$$

where $\hat{x}^{(0)}(1) = x^{(0)}(1)$, $k = 1,2,\cdots,m,\cdots$.

$\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \cdots, \hat{x}^{(0)}(m)$ are called the fitted values, and $\hat{x}^{(0)}(m+1), \hat{x}^{(0)}(m+2), \cdots \hat{x}^{(0)}(m+h)$ are called the predicted values.

2.2 Taylor Approximation Method in Grey Model

In this paper, Taylor approximation method in grey model (abbreviated as T-GM(1,1)), which is described as follows.

Algorithm of T-GM(1,1)

Step 1: Initialization

(a) Setting the updated times *K*. In this study, *K*=100 is used for T-GM(1,1).

(b) Setting objective function vector:

$$G = [x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)]^T \tag{11}$$

where $\left\{x^{(0)}(i), i = 1,2,\cdots,n\right\}$ are the measured data.

(c) Setting approximation function vector $F^{(K)}$:

$$F^{(K)} = [\hat{x}^{(0)(K)}(1), \hat{x}^{(0)(K)}(2), \cdots, \hat{x}^{(0)(K)}(n)]^T \tag{12}$$

where $\left\{\hat{x}_{(i)}^{(K)}, i = 1,2,\cdots,n\right\}$ are *K* times generated predicted series of indirect measurement model based on GM(1,1). When *K*= 0, $F^{(0)}$ is the predicted data series $\hat{x}^{(0)}$.

(d) Setting approximation parameters:

$$\hat{a}^{(K)} = [a,b]^T \tag{13}$$

where $\hat{a}^{(K)}$ is the *K*-th time of generated parameter series, $\hat{a}^{(0)}$ is the initial series of coefficients a and *b* of GM(1,1).

Step 2: Updating the calculation of approximated function vector $F^{(K+1)}$ according to the first order Taylor development:

$$F^{(K+1)} = F^{(K)} + F_a^{(K)}[a^{(K+1)} - a^{(K)}] + F_b^{(K)}[b^{(K+1)} - b^{(K)}] \quad (14)$$

$$F_a^{(K)} = \frac{\partial F^{(K)}}{\partial a^{(K)}} \approx \frac{F^{(K)}(a^{(K)} + C_a^{(K)}) - F^{(K)}(a^{(K)})}{C_a^{(K)}} \quad (15)$$

$$F_b^{(K)} = \frac{\partial F^{(K)}}{\partial b^{(K)}} \approx \frac{F^{(K)}(b^{(K)} + C_b^{(K)}) - F^{(K)}(b^{(K)})}{C_b^{(K)}} \quad (16)$$

$C_a^{(K)} = \frac{a^{(K)}}{h}$, $C_b^{(K)} = \frac{b^{(K)}}{h}$. Coefficient *h* is called the step length.

In this study, *h*=500 is used for T-GM(1,1).

Step 3: Setting the evaluation function $Q^{(K)}$

$$Q^{(K)} = [F_D^{(K)} - F_a^{(K)}\eta_a^{(K)} - F_b^{(K)}\eta_b^{(K)}]^T \cdot [F_D^{(K)} - F_a^{(K)}\eta_a^{(K)} - F_b^{(K)}\eta_b^{(K)}] \quad (17)$$

$$F_D^{(K)} = G - F^{(K)} \quad (18)$$

$$\eta^{(K)} = \begin{bmatrix} \eta_a^{(K)} \\ \eta_b^{(K)} \end{bmatrix} = \begin{bmatrix} \eta_a^{(K+1)} - \eta_a^{(K)} \\ \eta_b^{(K+1)} - \eta_b^{(K)} \end{bmatrix} \quad (19)$$

Step 4: Detecting the stop criterion

If $Q^{(K)} \le \varepsilon$ or *K*=100 for T-GM(1,1); stop; otherwise, go to Step 5. Where ε is the tolerance error.

Step 5: Updating the approximated parameters $\hat{a}^{(K)}$

$$\text{In order to minimize: } Q^{(K)} \to 0 \quad (20)$$

$$\text{let } \frac{\partial Q^{(K)}}{\partial \eta_a^{(K)}} = 0, \quad \frac{\partial Q^{(K)}}{\partial \eta_b^{(K)}} = 0 \quad (21)$$

and using Equation (17), the updated equation of parameters $\hat{a}^{(K)}$ can be obtained by

$$\hat{a}^{(K+1)} = \hat{a}^{(K)} + \frac{1}{H}[A^{(K)T}A^{(K)}]^{-1}A^{(K)T}F_D^{(K)} \quad (22)$$

$$A^{(K)} = [F_a^{(K)}, F_b^{(K)}] \quad (23)$$

H is adjustment coefficient. In this study, *H*=20 is used for T-GM(1,1).

Step 6: Increasing the updated times: *K*=*K*+1; go to Step 2.

End of algorithm

Using the optimization process, the parameters $\hat{a}^{(K)}$ are updated for *K* times, and the evaluation function $Q^{(K)}$ as the convergent error is reduced. When *K*=200, the researchers can find the optimal parameters and the convergent error is reduced to a minimum in this study. At this time, vector $F^{(K)}$ becomes the *K*-th predicted series $\hat{x}^{(0)(K)}(i)$ as the result of approximated calculation.

2.5 Error Analysis

In this study, researchers used mean absolute percentage error (MAPE), which can be calculated using Equation (24) as the error analysis method (Li et al., 2007; Zeng et al., 2010; Lin et al., 2013;

Nguyen et al., 2014; Nguyen et al., 2017). If the MAPE is less than 10%, the prediction result will be accepted (Nagai and Yamaguchi, 2004; Nguyen et al., 2019).

$$\text{MAPE} = \frac{1}{n}\sum_{k=2}^{n}\left|\frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)}\right| \times 100\% \tag{24}$$

3. BUILDING THE MATLAB TOOLBOX

3.1 Software Specifications and Requirements

(1) Windows 10.
(2) Screen resolution 1280×800.
(3) MATLAB R2019a version or upgrade versions.

This paper presents a sample program that is developed by MATLAB software, including many scientific functions due to the supply an experimental environment on the computer, and then a reliable program can be developed. The program for the above prediction model has been developed by MATLAB software (Wen and Chang 2005; Sheu et al., 2013; Nguyen et al., 2014; Nguyen et al., 2017; Nguyen et al., 2019).

3.2 The Program for Taylor Approximation Method in Grey Model

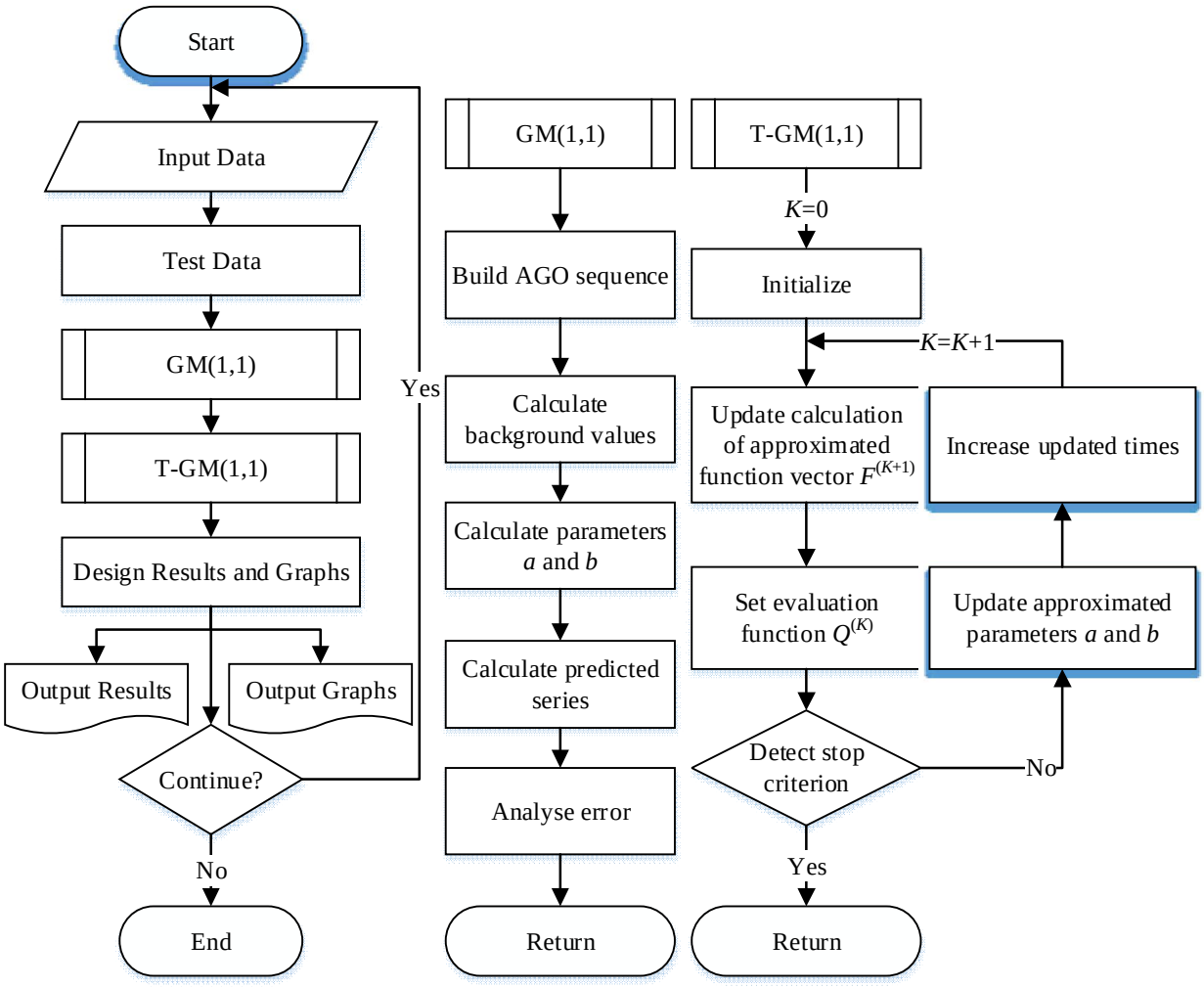

Figure 1. The flowchart for Taylor approximation method in grey model

The data processing cycle consists of six steps as follows.

Step 1: Input data. Data are the number of teachers and students at the ethnic minority boarding schools in the Mekong Delta region, Vietnam. The data have to be numerical, and written in *.csv file or *.xlsx file.

Step 2: Test data. The initial data have to be tested based on Equation (1) whether data are consistent with GM(1,1).

Step 3: Calculate parameters *a* and *b*; then calculate the predicted series ($\hat{x}^{(0)}(k)$) and error analysis for GM(1,1).

Step 4: Calculate parameters *a* and *b*; calculate the predicted series ($\hat{x}^{(0)}(k)$) and error analysis for Taylor approximation method in grey model.

Step 5: Design the results and the graphs to display the results and the graphs on a graphical user interface (GUI) visually. The user can save the results as an excel file and the graphs as an image file (JPG).

Step 6: Continue or exit program. If the user input a new data, the program will continue and back to step 1, or else the program will be closed.

4. RESULTS AND DISCUSSION

In this paper, data were collected at Department of Education and Training of Can Tho city and eight provinces of Vietnam in the Mekong Delta region (Kien Giang, An Giang, Tra Vinh, Soc Trang, Vinh Long, Hau Giang, Bac Lieu, and Ca Mau). Data are the number of teachers and students at the ethnic minority boarding schools in the Mekong Delta region, Vietnam from the 2005-2006 academic year to the 2019-2020 academic year (The data shown in Table 1).

Table 1. The number of teachers and students at the ethnic minority boarding schools

Unit: Person

Year	2015 - 2016	2016 - 2017	2017 - 2018	2018 - 2019	2019 - 2020
Number of Teachers	886	918	909	938	933
Number of Students	8842	9488	9635	10219	10436

Before using grey model, the initial data tested whether data are consistent with the prediction model. Test results showed that the data are consistent with the use of grey model to predict the number of teachers and students at the ethnic minority boarding schools in the Mekong Delta region, Vietnam (Test results shown in Fig. 2 and Fig. 3).

4.1 The prediction results for the number of teachers

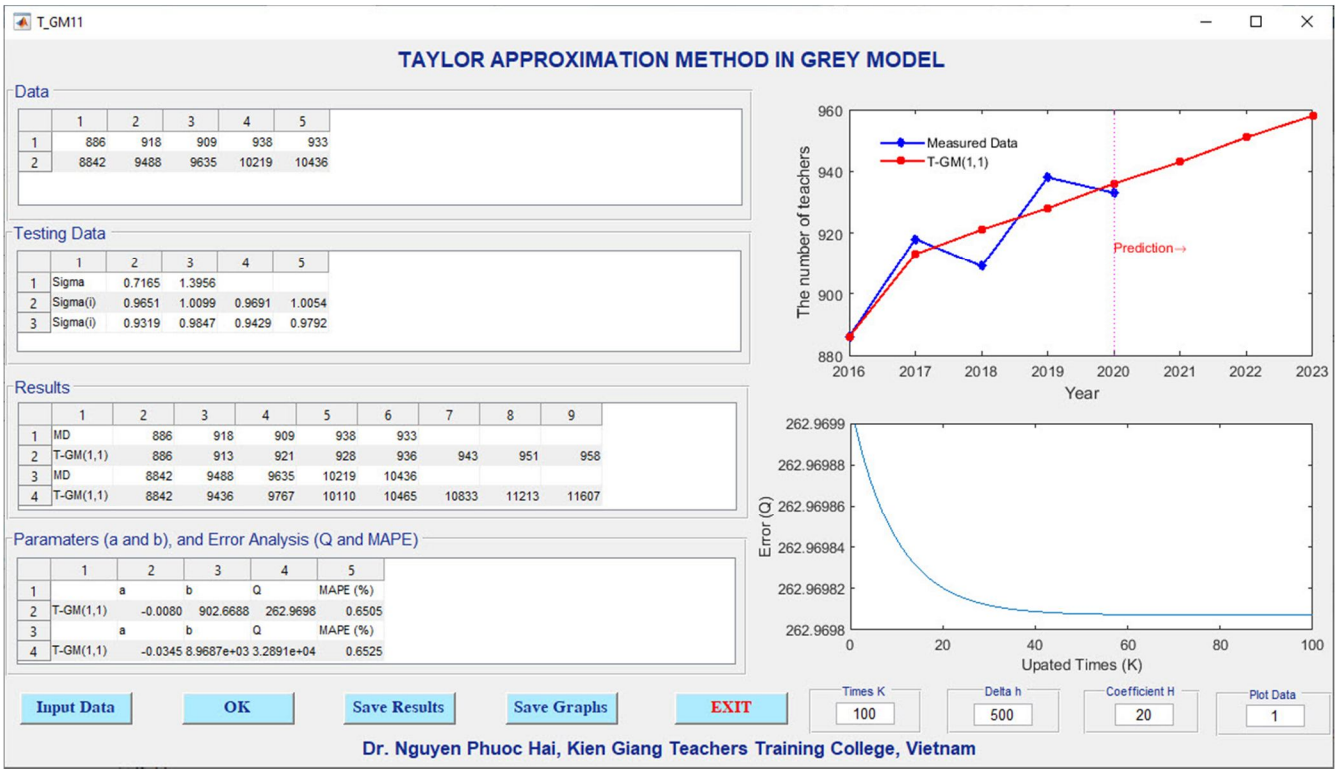

Figure 2. The GUI of the toolbox for the prediction of the number of teachers
(*x value represents school year and y value represents the number of teachers*)

According to the prediction results showed that T-GM(1,1) has high predictable power for the prediction of the number of teachers at the ethnic minority boarding schools in the Mekong Delta region, Vietnam. Establishing the original sequence for the number of teachers: $x^{(0)} = (886, 918, 909, 938, 933)$. Using T-GM(1,1) to calculate parameters *a* and *b*, the results obtained *a* = -0.0080 and *b* = 902.6688; the predicted values $\hat{x}^{(0)} = (886, 913, 921, 928, 936, 943, 951, 958)$; and the predicted error of T-GM(1,1): $Q = 262.9698$, $\text{MAPE} = 0.6505\%$.

4.2 The prediction results for the number of students

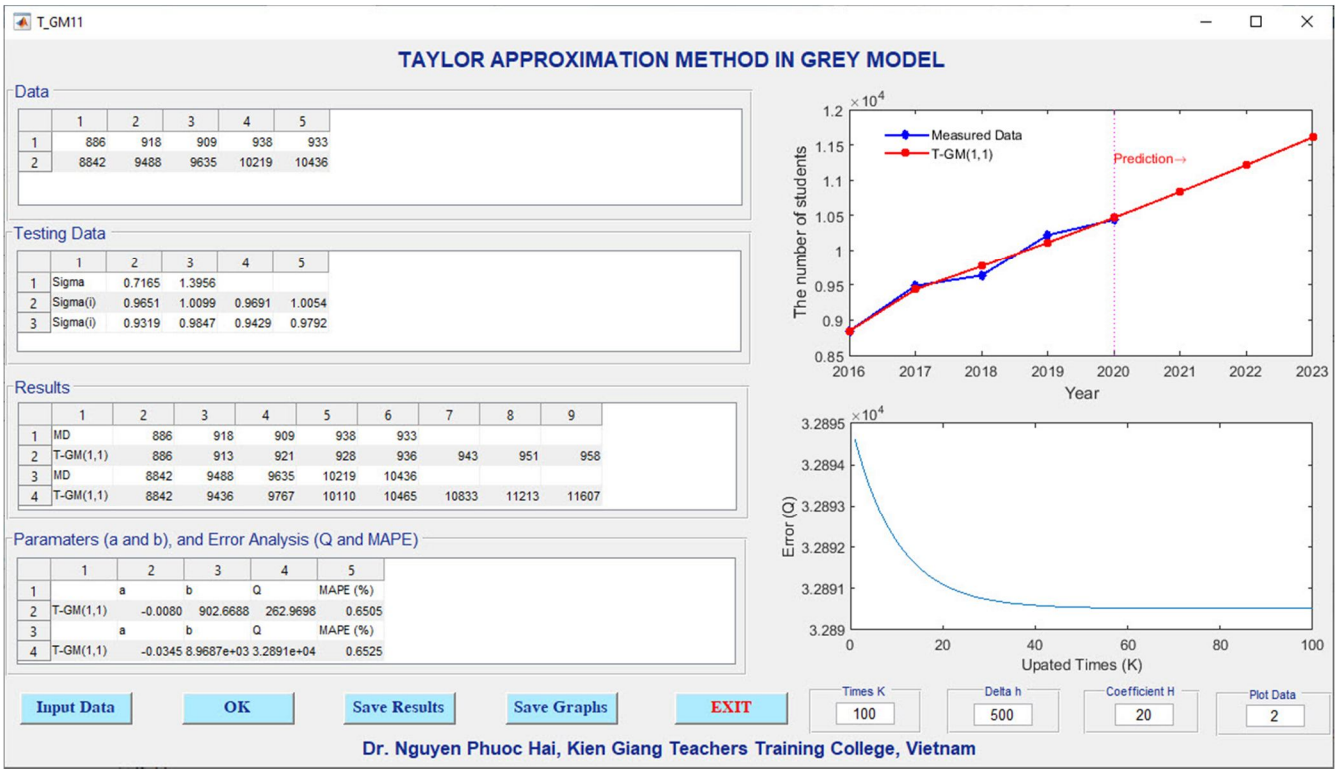

Figure 3. The GUI of the toolbox for the prediction of the number of students
(*x value represents school year and y value represents the number of students*)

According to the prediction results showed that T-GM(1,1) has high predictable power for the prediction of the number of students at the ethnic minority boarding schools in the Mekong Delta region, Vietnam. Establishing the original sequence for the number of students: $x^{(0)} = (8842, 9488, 9635, 10219, 10436)$. Using T-GM(1,1) to calculate parameters *a* and *b*, the results obtained *a* = -0.0345 and *b* = 8.97E+03; the predicted values $\hat{x}^{(0)} = (8842, 9436, 9767, 10110, 10465, 10833, 11213, 11607)$; and the predicted error of T-GM(1,1): $Q = 3.2891E+04$, $\text{MAPE} = 0.6525\%$.

According to the results showed that T-GM(1,1) is consistent with predicting the number of teachers and students at the ethnic minority boarding schools in the Mekong Delta region, Vietnam. In addition, the results indicated that the MATLAB toolbox can help to process data quickly, accurately, which displays the results and the graphs on the graphical user interface to predict the number of teachers and students at the ethnic minority boarding schools in the Mekong Delta region, Vietnam. These MAPE values of T-GM(1,1) for predicting the number of teachers and students are less than 10% (0.6505 and 0.6525). These results not only provide important information for educational administrators in Vietnam, but also help educational administrators to build education development strategies in the future. The predicted data has reliable information and high accuracy to contribute to the success in the educational development of the country.

Table 2. The results and the accuracy of Taylor approximation method in grey model

Unit: Person

Year	Measured Data		Prediction	
	Number of Teachers	**Number of Students**	**Number of Teachers**	**Number of Students**
2015-2016	886	8842	886	8842
2016-2017	918	9488	913	9436
2017-2018	909	9635	921	9767
2018-2019	938	10219	928	10110
2019-2020	933	10436	936	10465
2020-2021	-	-	943	10833
2021-2022	-	-	951	11213
2022-2023	-	-	958	11607
Q	-	-	262.9698	3.29E+04
MAPE(%)	-	-	0.6505	0.6525

5. CONCLUSIONS

Based on the findings from this study, some conclusions and suggestions are as follows:

(1) This study has successfully used Taylor approximation method in grey model (T-GM(1,1)) to predict the number of teachers and students at the ethnic minority boarding schools in the Mekong Delta region, Vietnam. The prediction results will provide important information for educational administrators to propose an appropriate policy for building education development strategies that are appropriate for the new conditions.

(2) This prediction model is not only a good method for predicting the number of teachers and students at the ethnic minority boarding schools in the Mekong Delta region, Vietnam, but also using in many fields such as educational measurement and statistics, medicine, business, and industry.

To sum up, this prediction model is actually useful for the prediction problems of uncertainty systems when the number of data is not enough for mathematical statistics, and this MATLAB toolbox not only helps user to process data quickly and accurately, but also displays the results on a graphical user interface visually.

REFERENCES

Chen, B., Sun, S. Q., and Liu, G. (2012), *An Optimized Unbiased GM (1, 1) Power Model for Forecasting MRO Spare Parts Inventory*, Modern Applied Science, 6, 12-17.

Chen, C. I., 2008, *Application of the novel nonlinear grey Bernoulli model for forecasting unemployment rate*, Chaos, Solitons and Fractals, 37, 278–287.

Deng, J. L. (1989), *Introduction to Grey System Theory*, Journal of Grey System, 1, 1-24.

Deng, L. R., Hu, Z. Y., Yang, S. G., and Yan, Y. Y. (2012), *Improved Non-equidistant Grey Model GM(1,1) Applied to the Stock Market*, Journal of Grey System, 15, 189-194.

Doan, D. H. (2005), *Moral education or political education in the Vietnamese educational system?*, Journal of Moral Education, 34, 451-463.

Li, G. D., Yamaguchi, D., and Nagai, M. (2005), *New Methods and Accuracy Improvement of GM According to Laplace Transform*, Journal of Grey System, 8, 13-24.

Li, G. D., Yamaguchi, D., Mizutani, K., and Nagai, M. (2007), *New Proposal and Accuracy Evaluation of Grey Prediction GM*, IEICE TRANS. FUNDAMENTALS, E90, 1188-1197.

Lin, Y. H., Chiu, C. C., Lin, Y. J., and Lee, P. C. (2013), *Rainfall Prediction Using Innovative Grey Model with The Dynamic Index*, Journal of Marine Science and Technology, 21, 63-75.

Nagai, M., and Yamaguchi, D. (2004), *Grey Theory and Engineering Application Method*. Tokyo: Kyoritsu Publisher.

Nguyen, P. H., Sheu, T. W., Nguyen, P. T., Pham, D H., & Nagai, M. (2014), *Taylor Approximation Method in Grey System Theory and Its Application to Predict the Number of Teachers and Students for Admission*, International Journal of Innovation and Scientific Research, *10*(2), 353-363.

Nguyen, P. H., Nguyen P. T., Ho C. P., Trinh T. K. B., & Nagai M. (2017), *The Prediction of the Admission Teacher's Number in Taiwan by using T-GM(1,n) and T-GM(2,n) Method*, Journal of Grey System, 20(*3*), 139-150.

Nguyen, P. H., Trinh T. K. B., & Nguyen P. T. (2019), *A New Prediction Model T-DGM(2,n) and Its Application*, International Journal of Uncertainty and Innovation Research, *1*(5), pp. 63-76.

Pai, T. Y., Ho, C. L., Chen, S. W., Lo, H. M., Sung, P. J., Lin, S. W., Lai, W. J., Tseng S. C., Ciou, S. P., Kuo, J. K., and Kao, J. Y. (2011), *Using Seven Types of GM (1, 1) Model to Forecast Hourly Particulate Matter Concentration in Banciao City of Taiwan*, Water, Air, & Soil Pollution, 217, 25-33.

Sheu, T. W., Nguyen, P. H., Nguyen, P. T., and Pham, D. H. (2013), *A Matlab Toolbox for AHP and LGRA-AHP to Analyze and Evaluate Factors in Making the Decision*, International Journal of Kansei Information, 4, 149-158.

Sheu, T. W., Nguyen, P. H., Nguyen, P. T., Pham, D. H., Tsai, C. P., & Nagai, M. (2014), *Using GM(2,1) and T-GM(2,1) to predict the number of students for admission*, Journal of Information and Computational Science, 11(*17*), 6085-6096.

Wen, K. L. and Chang, T. C. (2005), *The Research and Development of Completed GM(1,1) Model Toolbox Using Matlab*, International Journal of Computational Cognition, 3, 42-48.

Zeng, B., Liu, S. F., and Xie, N. M. (2010), *Prediction model of interval grey number based on DGM(1,1)*, Journal of Systems Engineering and Electronics, 21, 598–603.