

# Dự báo kết quả học tập của học sinh dựa trên sự kết hợp phương pháp gần đúng Taylor và các mô hình xám

Nguyễn Phước Hải<sup>\*,1</sup>, Tian-Wei Sheu<sup>2</sup>, Masatake Nagai<sup>2</sup>

<sup>1</sup>Trường Cao đẳng Sư phạm Kiên Giang,

Số 449, Đường Nguyễn Chí Thanh, Tp. Rạch Giá, tỉnh Kiên Giang

<sup>2</sup>Graduate Institute of Educational Information and Measurement,

National Taichung University of Education, Taiwan,

No. 140, Minsheng Rd., West Dist., Taichung City 40306, Taiwan (R.O.C.)

Nhận ngày 22 tháng 4 năm 2015

Chỉnh sửa ngày 29 tháng 5 năm 2015; chấp nhận đăng ngày 22 tháng 6 năm 2015

**Tóm tắt:** Mục đích của nghiên cứu này là dự báo kết quả học tập của học sinh dựa trên sự kết hợp phương pháp gần đúng Taylor với hai mô hình xám GM(1,1) và GM(2,1). Hai mô hình kết hợp T-GM(1,1) và T-GM(2,1) có thể đạt được các giá trị dự báo tối ưu nhất bằng cách tính gần đúng nhiều lần để cải thiện độ chính xác dự báo của hai mô hình xám. Ngoài ra, người nghiên cứu đã sử dụng phần mềm MATLAB để thiết kế một hộp công cụ MATLAB cho hai mô hình kết hợp này. Kết quả nghiên cứu này sẽ cung cấp thông tin rất quan trọng cho giáo viên và cán bộ quản lý giáo dục giúp cho họ tuyển chọn học sinh có quá trình học tập ổn định để bồi dưỡng học sinh giỏi, đồng thời cải thiện kết quả học tập đối với học sinh có quá trình học tập không ổn định nhằm đáp ứng các yêu cầu và mục tiêu của giáo dục.

**Từ khóa:** Kết quả học tập, phương pháp gần đúng Taylor, mô hình xám, hộp công cụ MATLAB, quá trình học tập.

## 1. Giới thiệu

Dự báo phát triển giáo dục là vấn đề có ý nghĩa quan trọng nhằm tạo ra cơ sở khoa học cho hoạch định chính sách, chiến lược phát triển giáo dục. Dự báo trong giáo dục ngày càng có vai trò và nhiệm vụ quan trọng trong việc xây dựng chiến lược phát triển giáo dục đúng hướng, hợp quy luật, xu thế và xứng tầm với thời đại. Dự báo dựa trên mô hình là một cách tiếp cận những thông tin cho tương

lai bằng công cụ mô hình hóa. Thông qua việc mô phỏng lại quá khứ và so sánh các giá trị dự báo được tính toán bằng mô hình với dữ liệu thực tế, nếu sai số nằm trong giới hạn cho phép thì mô hình đó được coi là có thể áp dụng được. Trong bài viết này, người nghiên cứu dự báo kết quả học tập (KQHT) của học sinh (HS) dựa trên sự kết hợp phương pháp gần đúng Taylor với hai mô hình xám GM(1,1) và GM(2,1) (viết tắt là T-GM(1,1) và T-GM(2,1)). Kết quả nghiên cứu sẽ cung cấp thông tin quan trọng cho giáo viên (GV) và cán bộ quản lý giáo dục, giúp cho họ chủ động phân loại HS, sắp xếp lớp học hợp lý,

\* Tác giả liên hệ. ĐT: 84-918588970  
Email: phuochai1979@gmail.com

tuyển chọn HS có quá trình học tập ổn định để bồi dưỡng HS giỏi, đồng thời cải thiện KQHT đối với HS có quá trình học tập không ổn định nhằm đáp ứng các yêu cầu và mục tiêu của giáo dục.

Năm 1982, Deng đã đề xuất lí thuyết hệ thống xám (Grey System Theory). Lí thuyết hệ thống xám nghiên cứu hệ thống thông tin không chắc chắn với số liệu có cỡ mẫu nhỏ và hệ thống thông tin không đầy đủ [1]. Trong những năm gần đây, lí thuyết hệ thống xám đã trở thành một phương pháp rất hiệu quả để giải quyết vấn đề đối với các dữ liệu rời rạc và không đầy đủ thông tin [2]. Mô hình xám dựa trên lí thuyết hệ thống xám là mô hình dự báo đã được ứng dụng trong nhiều lĩnh vực khác nhau [3-5]. Mô hình xám GM(1,1) (Grey Model (1, 1)) là một trong những phần quan trọng trong lí thuyết hệ thống xám và được xem là cốt lõi của mô hình dự báo xám [6]. Ưu điểm của mô hình này là có thể sử dụng khi số lượng dữ liệu không đủ để thực hiện các phương pháp phân tích thống kê. Nó chỉ cần một lượng nhỏ dữ liệu và dữ liệu mẫu ngẫu nhiên là có thể tính toán và đưa ra kết quả dự báo [7, 8]. Trong những năm gần đây, mô hình xám đã được áp dụng rộng rãi trong nhiều lĩnh vực nghiên cứu để giải quyết hiệu quả các vấn đề dự báo của các hệ thống không chắc chắn [4, 9]. Hiện nay, lí thuyết hệ thống xám nói chung và mô hình xám nói riêng vẫn chưa được sử dụng phổ biến ở Việt Nam, đặc biệt là dùng để dự báo trong lĩnh vực giáo dục.

Bên cạnh đó, nhiều nhà nghiên cứu cũng đã chỉ ra rằng độ chính xác dự báo của mô hình xám là chưa cao [10-12]. Các tham số của mô hình xám chưa phải là các tham số tối ưu. Vì vậy, nhiều nhà nghiên cứu đã sử dụng nhiều phương pháp khác nhau để cải thiện độ chính xác của mô hình xám [5, 13-15]. Trong

thời gian gần đây, sự kết hợp phương pháp gần đúng Taylor và mô hình xám GM(1,1) đã được sử dụng để dự báo kết quả học tập của học sinh ở Đài Loan [16], và sự kết hợp phương pháp gần đúng Taylor và mô hình xám GM(2,1) cũng được sử dụng để dự báo số lượng học sinh nhập học ở Đài Loan [17], phương pháp này đã cải thiện đáng kể độ chính xác của các mô hình dự báo. Tuy nhiên khi sử dụng một trong hai mô hình kết hợp T-GM(1,1) và T-GM(2,1) để dự báo thì độ chính xác có thể chưa cao. Bởi vì có những dữ liệu chỉ phù hợp với một trong hai mô hình kết hợp. Vì vậy, trong nghiên cứu này người nghiên cứu sử dụng kết hợp phương pháp gần đúng Taylor với hai mô hình xám GM(1,1) và GM(2,1) để điều chỉnh tối ưu các tham số, làm cho sai số của hai mô hình xám GM(1,1) và GM(2,1) giảm đến mức tối thiểu. Hơn nữa, người nghiên cứu sử dụng phần mềm MATLAB để thiết kế một hộp công cụ MATLAB cho hai mô hình dự báo này. Hộp công cụ MATLAB giúp cho quá trình tính toán dễ dàng, nhanh chóng, chính xác, hiển thị kết quả và hình ảnh trên giao diện đồ họa người dùng một cách trực quan sinh động.

## 2. Cơ sở lý thuyết và phương pháp nghiên cứu

### 2.1. Mô hình xám GM(1,1)

Trước khi sử dụng mô hình xám GM(1,1) dữ liệu ban đầu cần phải kiểm định theo công thức sau [14]:

$$\sigma(i) = \frac{x^{(0)}(i-1)}{x^{(0)}(i)}, i = 2, 3, \dots, n. \quad (1)$$

Nếu tất cả giá trị  $\sigma(i)$  đều nằm trong khoảng giá trị  $\sigma^{(0)}(i) = \left( e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}} \right)$  thì có

thể sử dụng mô hình xám GM(1,1) để dự báo. Nếu không thỏa mãn điều kiện trên thì phải sử dụng một mô hình xám khác để dự báo.

Mô hình GM(1,1) được tính dựa trên phương trình vi phân sau đây [1]:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b. \quad (2)$$

*Bước 1:* Tính các giá trị  $x^{(1)}$  bằng cách sử dụng phương pháp cộng tích lũy:

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)). \quad (3)$$

$$x^{(1)} = (x^{(0)}(1), x^{(0)}(1) + x^{(0)}(2), \dots, x^{(0)}(1) + \dots + x^{(0)}(n)). \quad (4)$$

$$x^{(1)} = \left( \sum_{k=1}^1 x^{(0)}(k), \sum_{k=1}^2 x^{(0)}(k), \dots, \sum_{k=1}^n x^{(0)}(k) \right). \quad (5)$$

*Bước 2:* Thiết lập phương trình của mô hình xám GM(1,1) và tính các giá trị  $z^{(1)}$

$$x^{(0)}(k) + az^{(1)}(k) = b. \quad (6)$$

$$\text{Trong đó } z^{(1)}(k) = 0.5x_1^{(1)}(k) + 0.5x_1^{(1)}(k-1), \quad k = 2, 3, \dots, n. \quad (7)$$

*Bước 3:* Tính các tham số  $a$  và  $b$

Tham số  $a$  và  $b$  của mô hình xám GM(1,1) được tính dựa trên phương pháp bình phương tối thiểu, cụ thể như sau:

$$\hat{a} = \begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y. \quad (8)$$

$$\text{Trong đó, } B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}, \quad Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}. \quad (9)$$

*Bước 4:* Thiết lập công thức để tính các giá trị dự báo của mô hình

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

Sau đó tính được các giá trị dự báo của mô hình xám GM(1,1) dựa trên công thức sau:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), \quad k = 1, 2, \dots, n, n+1, n+2, \dots \quad (11)$$

Trong đó  $\hat{x}^{(0)}(1) = x^{(0)}(1)$ .

Trong đó,  $a$  và  $b$  là các hệ số.

Dữ liệu ban đầu được xem là một chuỗi giá trị  $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$ , trong đó  $n \geq 4$ . Trong nghiên cứu này  $x^{(0)}$  là KQHT của HS được thống kê trong ba năm học. Dữ liệu sau khi được kiểm định sẽ được tính toán theo các bước sau đây:

## 2.2. Mô hình xám GM(2,1)

Giả sử rằng  $x^{(0)}$  là chuỗi giá trị ban đầu của mô hình xám GM(2,1) gồm có  $n$  giá trị.

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)). \quad (12)$$

Tính các giá trị  $x^{(1)}$  bằng phương pháp cộng tích lũy từ  $x^{(0)}$

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)). \quad (13)$$

Trong đó  $x^{(1)}(1) = x^{(0)}(1)$ , và

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), \quad k = 1, 2, \dots, n. \quad (14)$$

Phương trình vi phân của mô hình xám GM(2,1) như sau:

$$\begin{cases} \hat{x}^{(1)}(1) = x^{(0)}(1), \quad (\hat{x}^{(1)}(t))'_{t=1} = \frac{1}{2}(x^{(0)}(3) - x^{(0)}(1)) \\ \frac{d^2 x^{(1)}}{dt^2} + a_1 \frac{dx^{(1)}}{dt} + a_2 x^{(1)} = b \end{cases}. \quad (15)$$

Trong đó  $\hat{x}^{(1)}(1)$  và  $(\hat{x}^{(1)}(t))'_{t=1}$  là giá trị của hệ thống tại thời điểm ban đầu. Nó có thể cho thấy rằng giải pháp cho  $\hat{x}^{(1)}(k)$  là

$$\hat{x}^{(1)}(k) = \hat{x}_*^{(1)}(k) + \frac{b}{a_2}. \quad (16)$$

Trong đó  $\hat{x}_*^{(1)}(k)$  được gọi là giải pháp chung cho phương trình vi phân sau đây.

$$\frac{d^2 x^{(1)}}{dt^2} + a_1 \frac{dx^{(1)}}{dt} + a_2 x^{(1)} = b. \quad (17)$$

Dựa theo mối quan hệ giữa  $a_1$  và  $a_2$  cho thấy có ba giải pháp cho phương trình (17) [18]. Tuy nhiên trong nghiên cứu này, mô hình xám GM(2,1) được tính như sau:

$$\hat{x}^{(1)}(k+1) = C_1 e^{\lambda_1 k} + C_2 e^{\lambda_2 k} + \frac{b}{a_2}. \quad (18)$$

$$\text{Trong đó } \lambda_1 = \frac{-a_1 + \sqrt{a_1^2 - 4a_2}}{2}. \quad (19)$$

$$\lambda_2 = \frac{-a_1 - \sqrt{a_1^2 - 4a_2}}{2}. \quad (20)$$

$$C_1 = \frac{1}{\lambda_2 - \lambda_1} \left[ \lambda_2 x^{(0)}(1) - \frac{1}{2} [x^{(0)}(3) - x^{(0)}(1)] - \lambda_2 \frac{b}{a_2} \right]. \quad (21)$$

$$C_2 = \frac{1}{\lambda_1 - \lambda_2} \left[ \lambda_1 x^{(0)}(1) - \frac{1}{2} [x^{(0)}(3) - x^{(0)}(1)] - \lambda_1 \frac{b}{a_2} \right]. \quad (22)$$

Các tham số  $a_1$ ,  $a_2$ , và  $b$  được tính như sau:

$$[a_1 \ a_2 \ b]^T = (B^T B)^{-1} B^T Y. \quad (23)$$

$$B = \begin{bmatrix} -x^{(0)}(2) & -z^{(1)}(2) & 1 \\ -x^{(0)}(3) & -z^{(1)}(3) & 1 \\ \vdots & \vdots & \vdots \\ -x^{(0)}(n) & -z^{(1)}(n) & 1 \end{bmatrix}. \quad (24)$$

$$Y = \begin{bmatrix} x^{(0)}(2) - x^{(0)}(1) \\ x^{(0)}(3) - x^{(0)}(2) \\ \vdots \\ x^{(0)}(n) - x^{(0)}(n-1) \end{bmatrix}. \quad (25)$$

$$\text{Trong đó } z^{(1)}(k) = 0.5x_1^{(1)}(k) + 0.5x_1^{(1)}(k-1), \quad k = 2, 3, \dots, n. \quad (26)$$

Sau đó tính được các giá trị dự báo của mô hình xám GM(2,1) dựa trên công thức sau:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k). \quad (27)$$

Trong đó  $\hat{x}^{(0)}(1) = x^{(0)}(1)$ . Dựa trên công thức (27), các giá trị  $\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \dots, \hat{x}^{(0)}(n)$  được cho là phù hợp với giá trị thực tế của mô hình xám GM(2,1), và  $\hat{x}^{(0)}(n+1), \hat{x}^{(0)}(n+2), \dots$  được gọi là các giá trị dự báo của mô hình xám GM(2,1).

### 2.3. Phương pháp gần đúng Taylor trong các mô hình xám

Trong bài viết này, phương pháp gần đúng Taylor được sử dụng kết hợp với hai mô hình xám GM(1,1) và GM(2,1) để làm tăng độ chính xác các giá trị dự báo. Thuật toán của hai mô hình T-GM(1,1) và T-GM(2,1) được mô tả như sau [14].

Thuật toán của hai mô hình T-GM(1,1) và T-GM(2,1)

*Bước 1:* Khởi tạo

(a) Thiết lập số lần cập nhật  $K$ . Trong nghiên cứu này  $K=100$  đã được sử dụng.

(b) Thiết lập các giá trị cần tối ưu hóa:

$$G = [x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)]^T. \quad (28)$$

Trong đó  $\{x^{(0)}(k), k = 1, 2, \dots, n\}$  là dữ liệu thực tế đo lường được.

(c) Thiết lập các giá trị gần đúng  $F^{(K)}$ :

$$F^{(K)} = [\hat{x}^{(0)(K)}(1), \hat{x}^{(0)(K)}(2), \dots, \hat{x}^{(0)(K)}(n)]^T. \quad (29)$$

Trong đó  $\{\hat{x}_{(k)}^{(K)}, k = 1, 2, \dots, n\}$  là chuỗi giá trị dự báo được tạo ra tương ứng với số lần cập nhật  $K$  dựa trên mô hình xám GM(1,1) hoặc GM(2,1). Khi  $K=0$ ,  $F^{(0)}$  là chuỗi giá trị dự báo  $\hat{x}^{(0)}$  của mô hình xám GM(1,1) hoặc GM(2,1).

(d) Thiết lập các tham số gần đúng của mô hình:

$$\hat{a}^{(K)} = [a_i, b]^T, \quad i = 1, 2. \quad (30)$$

Trong đó  $\hat{a}^{(K)}$  là các tham số được tạo ra tương ứng với số lần cập nhật  $K$ ,  $\hat{a}^{(0)}$  là các

tham số ban đầu  $a_1$  and  $b$  của mô hình xám GM(1,1), hoặc  $a_1, a_2$ , và  $b$  của mô hình xám GM(2,1).

*Bước 2:* Cập nhật tính các giá trị gần đúng  $F^{(K+1)}$  dựa theo tính toán khai triển Taylor cấp một của phương pháp gần đúng Taylor như sau:

$$F^{(K+1)} = F^{(K)} + F_{a_i}^{(K)}[a_i^{(K+1)} - a_i^{(K)}] + F_b^{(K)}[b^{(K+1)} - b^{(K)}]. \quad (31)$$

$$F_{a_i}^{(K)} = \frac{\partial F^{(K)}}{\partial a_i^{(K)}} \approx \frac{F^{(K)}(a_i^{(K)} + C_{a_i}^{(K)}) - F^{(K)}(a_i^{(K)})}{C_{a_i}^{(K)}}. \quad (32)$$

$$F_b^{(K)} = \frac{\partial F^{(K)}}{\partial b^{(K)}} \approx \frac{F^{(K)}(b^{(K)} + C_b^{(K)}) - F^{(K)}(b^{(K)})}{C_b^{(K)}}. \quad (33)$$

$C_{a_i}^{(K)} = \frac{a^{(K)}}{h}$ ,  $C_b^{(K)} = \frac{b^{(K)}}{h}$ . Hệ số  $h$  được gọi là độ dài bước tính toán. Trong nghiên cứu này,

$h=500$  đã được sử dụng.

*Bước 3:* Thiết lập đánh giá sai số  $Q^{(K)}$

$$Q^{(K)} = [F_D^{(K)} - F_{a_i}^{(K)}\eta_{a_i}^{(K)} - F_b^{(K)}\eta_b^{(K)}]^T \cdot [F_D^{(K)} - F_{a_i}^{(K)}\eta_{a_i}^{(K)} - F_b^{(K)}\eta_b^{(K)}]. \quad (34)$$

$$F_D^{(K)} = G - F^{(K)}. \quad (35)$$

$$\eta^{(K)} = \begin{bmatrix} \eta_{a_i}^{(K)} \\ \eta_b^{(K)} \end{bmatrix} = \begin{bmatrix} \eta_{a_i}^{(K+1)} - \eta_{a_i}^{(K)} \\ \eta_b^{(K+1)} - \eta_b^{(K)} \end{bmatrix}. \quad (36)$$

*Bước 4:* Xác định tiêu chí dừng quá trình tính toán

Nếu  $Q^{(K)} \leq \varepsilon$  hoặc  $K=100$ , quá trình tính toán sẽ dừng lại; ngược lại, quá trình sẽ tiếp tục đến bước 5. Trong đó  $\varepsilon$  là sai số chấp nhận ( $\varepsilon = 0,01$ ).

*Bước 5:* Cập nhật các tham số gần đúng  $\hat{a}^{(K)}$

Để cho sai số tiến gần đến 0:

$$Q^{(K)} \rightarrow 0. \quad (37)$$

$$\text{Khi đó } \frac{\partial Q^{(K)}}{\partial \eta_{a_i}^{(K)}} = 0, \quad \frac{\partial Q^{(K)}}{\partial \eta_b^{(K)}} = 0. \quad (38)$$

Sử dụng công thức (34) để đánh giá sai số và tính toán cập nhật các tham số,  $\hat{a}^{(K)}$  tiếp tục được tính dựa trên công thức sau:

$$\hat{a}^{(K+1)} = \hat{a}^{(K)} + \frac{1}{H} [A^{(K)T} A^{(K)}]^{-1} A^{(K)T} F_D^{(K)}. \quad (39)$$

$$A^{(K)} = [F_{a_i}^{(K)}, F_b^{(K)}]. \quad (40)$$

Trong đó  $H$  là hệ số điều chỉnh. Trong nghiên cứu này,  $H=20$  đã được sử dụng.

*Bước 6:* Tăng số lần cập nhật:  $K=K+1$ ; trở về bước 2.

*Kết thúc thuật toán*

Bằng cách sử dụng phương pháp gần đúng Taylor các tham số  $\hat{a}^{(K)}$  được cập nhật liên tục đến  $K$  lần, sai số  $Q^{(K)}$  giảm dần đến mức tối thiểu. Trong nghiên cứu này, khi  $K=100$ , người nghiên cứu có thể tìm thấy các tham số tối ưu và độ chính xác của dự báo tăng lên. Tại thời điểm này, vector  $F^{(K)}$  trở thành chuỗi giá trị dự báo và  $\hat{x}^{(0)(K)}(i)$  được

xem như là kết quả tính gần đúng dựa trên phương pháp gần đúng Taylor kết hợp với hai mô hình xám GM(1,1) hoặc GM(2,1).

2.4. Phân tích sai số

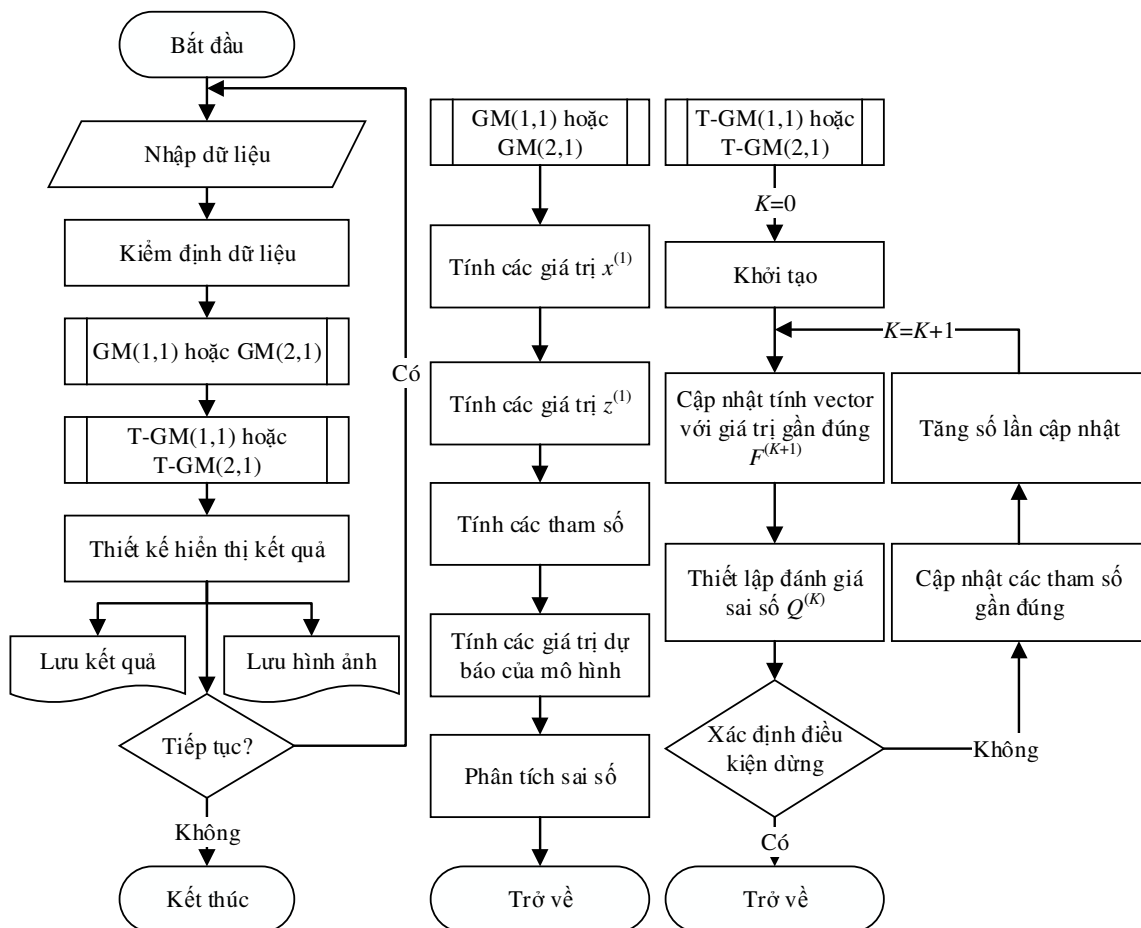
Trong nhiều nghiên cứu về mô hình dự báo, các nhà nghiên cứu thường sử dụng phần trăm sai số tuyệt đối trung bình (Mean Absolute Percentage Error, MAPE) để phân tích sai số dựa trên các giá trị dự báo của mô hình so với các giá trị thực tế để kiểm tra sự phù hợp của mô hình dự báo [19, 20].

$$MAPE = \left( \frac{1}{n} \sum_{k=1}^n \left| \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \right| \right) \times 100\% .(41)$$

Căn cứ một số nghiên cứu về việc sử dụng phần trăm sai số tuyệt đối trung bình cho thấy nếu  $MAPE < 10\%$  thì số liệu dự báo đạt yêu cầu khi sử dụng mô hình dự báo [18, 21, 22].

2.5. Thiết kế hộp công cụ MATLAB

Phần mềm MATLAB thường được sử dụng để thiết kế một hộp công cụ MATLAB trong quá trình tính toán phức tạp [23, 24]. Trong nghiên cứu này người nghiên cứu đã thiết kế một hộp công cụ MATLAB cho hai mô hình T-GM(1,1) và T-GM(2,1). Chương trình xử lý dữ liệu của hộp công cụ MATLAB được tóm tắt gồm có 6 bước như sau:



Hình 1. Lưu đồ của mô hình dự báo T-GM(1,1) và T-GM(2,1).

**Bước 1:** Nhập dữ liệu. Dữ liệu là KQHT của HS được mã hóa bằng số dưới dạng tập tin \*.csv hoặc \*.xlsx.

**Bước 2:** Kiểm định dữ liệu xem phù hợp với mô hình dự báo T-GM(1,1) hay T-GM(2,1).

**Bước 3:** Dựa trên mô hình xám để tính các tham số  $a$  và  $b$  đối với GM(1,1) hoặc  $a_1, a_2$  và  $b$  đối với GM(2,1); sau đó tính các giá trị dự báo  $\hat{x}^{(0)}(k)$  và phân tích sai số ( $Q, MAPE$ ).

**Bước 4:** Dựa trên mô hình T-GM(1,1) để tính các tham số  $a$  và  $b$  hoặc mô hình T-GM(2,1) để tính các tham số  $a_1, a_2$  và  $b$ ; sau đó tính các giá trị dự báo  $\hat{x}^{(0)}(k)$  và phân tích sai số ( $Q, MAPE$ ).

**Bước 5:** Thiết kế hiển thị các kết quả và hình ảnh trên giao diện đồ họa người dùng. Người sử dụng có thể lưu lại kết quả dưới dạng tập tin \*.csv hoặc \*.xlsx và hình ảnh dưới dạng tập tin \*.JPG.

**Bước 6:** Tiếp tục hoặc thoát khỏi chương trình. Nếu người sử dụng nhập dữ liệu mới vào chương trình sẽ được tiếp tục trở về bước 1, ngược lại chương trình sẽ đóng lại.

### 3. Kết quả nghiên cứu và thảo luận

#### 3.1. Kiểm định dữ liệu

Dữ liệu trong nghiên cứu này được lấy từ một trường THCS của huyện Giồng Riềng, tỉnh Kiên Giang. Dữ liệu là KQHT môn Sinh học của 30 HS trong ba năm học tương ứng với sáu học kỳ học tập từ lớp 6 đến lớp 8 (dữ liệu được trình bày ở Bảng 1). Trong bài báo này, người nghiên cứu sử dụng hai mô hình T-GM(1,1) và T-GM(2,1) để dự báo KQHT môn Sinh học của 30 HS ở học kỳ tiếp theo, sau đó so sánh kết quả dự báo với dữ liệu thực tế để kiểm tra độ chính xác của mô hình

dự báo. Trước khi tiến hành nghiên cứu, người nghiên cứu đã kiểm tra độ tin cậy của dữ liệu thông qua việc kiểm định hệ số Cronbach's Alpha. Hệ số Cronbach's Alpha của dữ liệu trong nghiên cứu này là 0,968, điều này cho thấy dữ liệu có độ tin cậy cao.

Trước khi sử dụng mô hình dự báo, dữ liệu được kiểm định dựa trên công thức (1) để xem dữ liệu phù hợp với mô hình dự báo T-GM(1,1) hay T-GM(2,1). Lưu đồ kiểm định dữ liệu để chọn mô hình dự báo được trình bày ở Hình 2. Trong nghiên cứu này có 22 số liệu có giá trị  $\sigma(i)$  nằm trong khoảng giá trị  $\sigma^{(0)}(i) = (0,75; 1,33)$  và có 8 số liệu không thỏa mãn điều kiện này. Đối với 8 số liệu không đạt khi kiểm tra dữ liệu người nghiên cứu sử dụng mô hình T-GM(2,1) để dự báo.

#### 3.2. Kết quả nghiên cứu

Trong bài viết này, dữ liệu ban đầu gồm có 30 số liệu tương ứng với KQHT môn Sinh học của 30 HS. Kết quả dự báo KQHT và sai số dựa trên hai mô hình T-GM(1,1) và T-GM(2,1) được trình bày ở Bảng 2. Sau đây là phần mô tả cách tính từng bước cho số liệu HS S1 dựa trên sự kết hợp giữa phương pháp gần đúng Taylor và mô hình xám GM(1,1).

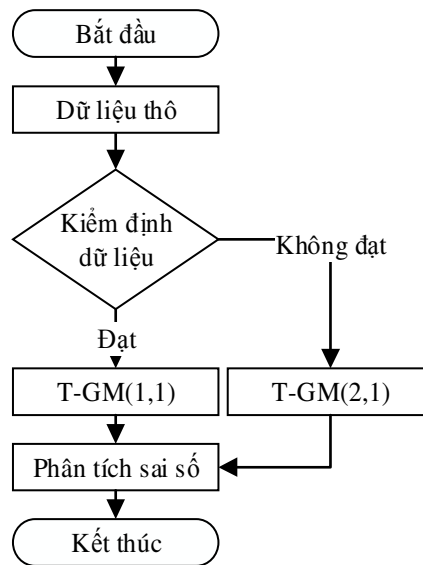
Số liệu thô của HS S1  $x^{(0)} = (8,6; 8,2; 8,0; 7,8; 7,6; 7,5)$ , áp dụng công thức (5) sẽ tính được  $x^{(1)} = (8,6; 16,8; 24,8; 32,6; 40,2; 47,7)$  và công thức (7) tính được  $z^{(1)} = (12,7; 20,8; 28,7; 36,4; 44,0)$ . Sau đó sử dụng công thức (8) sẽ tính được các tham số  $a$  và  $b$  ( $a = 0,0231$  và  $b = 8,4778$ ). Sau khi tính được  $a$  và  $b$  thì thay vào công thức (10) sẽ tính được các giá trị dự báo của mô hình GM(1,  $\hat{x}^{(0)} = (8,6; 8,2; 8,0; 7,8; 7,6; 7,5; 7,3)$ . Từ kết quả  $\hat{x}^{(0)}$  có thể thấy được KQHT của HS S1 dự báo cho học kỳ tiếp theo là 7,3. Sử



dùng kết quả dự báo so sánh với số liệu thực tế để phân tích sai số cho mô hình xám GM(1,1) dựa theo công thức (41), kết quả sai số MAPE = 0,2326%.

Tuy nhiên, khi sử dụng mô hình kết hợp T-GM(1,1) với các hệ số  $K=100$ ,  $h=500$  và  $H=20$ . Kết quả tính được các giá trị dự báo của mô hình T-GM(1,1)  $\hat{x}^{(0)} = (8,6; 8,2; 8,0; 7,8; 7,6; 7,5; 7,3)$ .

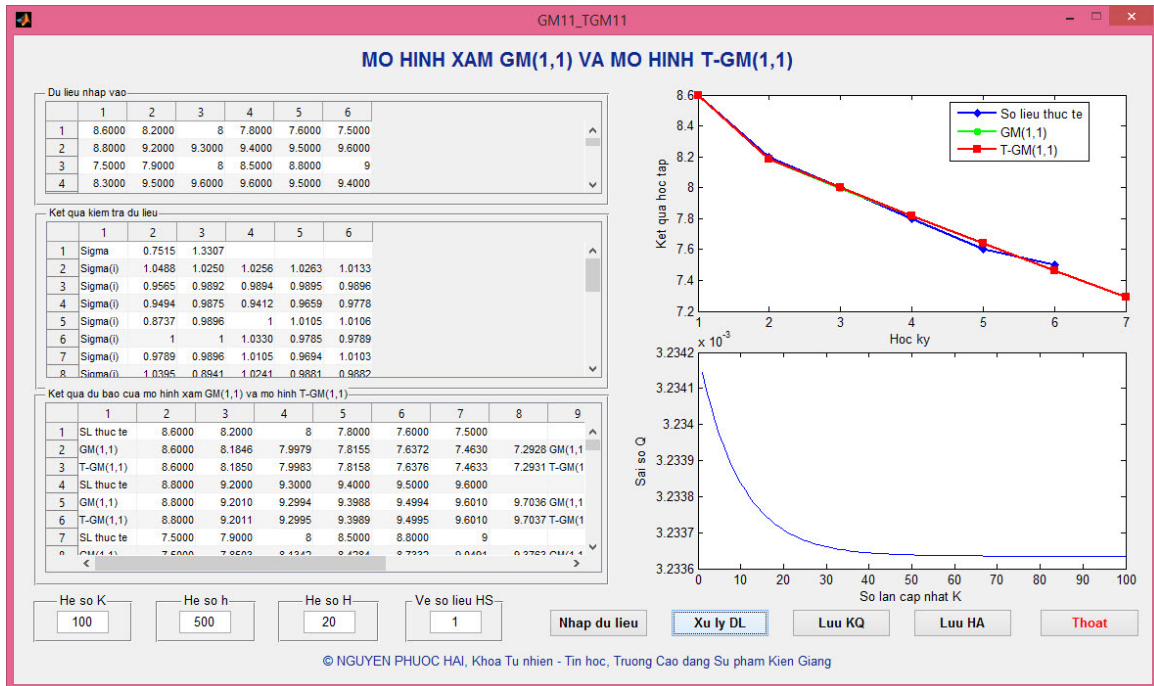
Từ kết quả  $\hat{x}^{(0)}$  có thể thấy được KQHT của HS S1 dự báo cho học kỳ tiếp theo là 7,3 và kết quả sai số MAPE = 0,2318%. Kết quả trên có thể thấy được trên giao diện đồ họa người dùng khi sử dụng hộp công cụ MATLAB để tính toán (Hình 3). Trên giao diện đồ họa này có thể thấy sai số  $Q$  của mô hình T-GM(1,1) được điều chỉnh giảm dần đến mức tối thiểu.



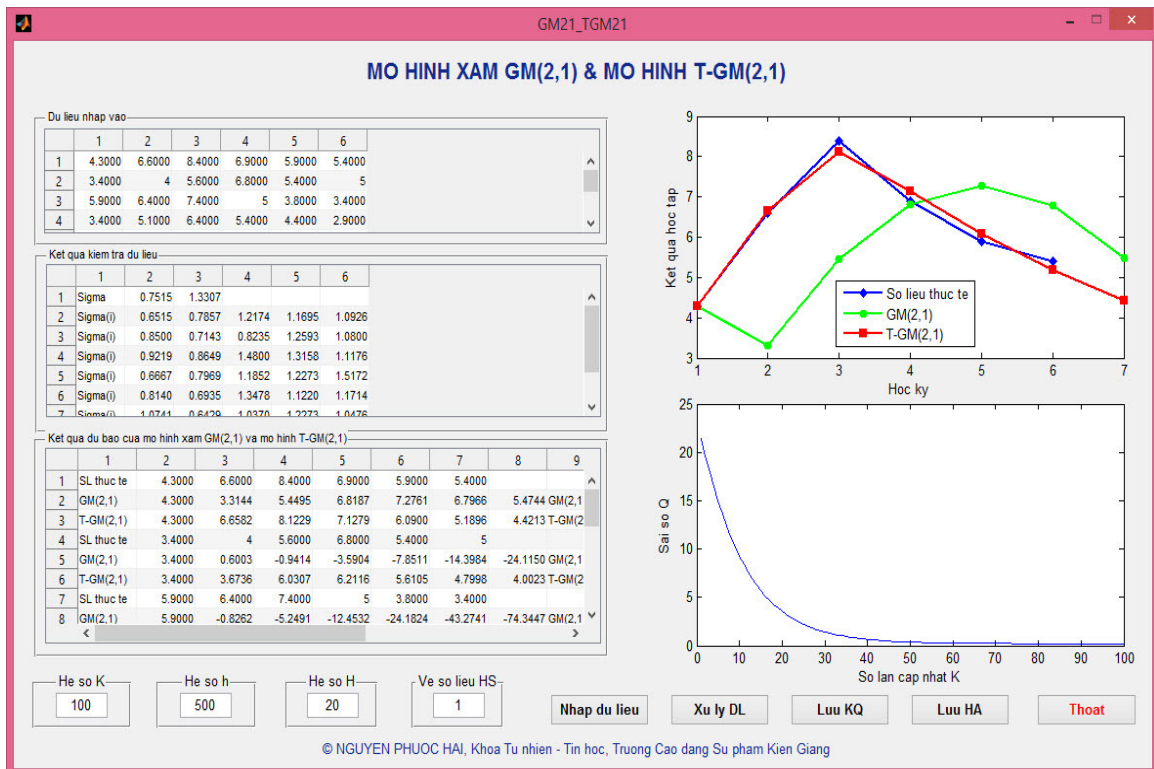
Hình 2. Lưu đồ kiểm định dữ liệu để chọn mô hình dự báo.

Bảng 1. Kết quả học tập môn Sinh học của 30 học sinh

Mã HS	Lớp 6		Lớp 7		Lớp 8		Mã HS	Lớp 6		Lớp 7		Lớp 8	
	HK1	HK2	HK1	HK2	HK1	HK2		HK1	HK2	HK1	HK2	HK1	HK2
S1	8,6	8,2	8,0	7,8	7,6	7,5	S16	5,8	5,4	8,4	8,1	6,6	6,3
S2	8,8	9,2	9,3	9,4	9,5	9,6	S17	8,0	8,2	8,5	8,9	8,6	8,8
S3	4,3	6,6	8,4	6,9	5,9	5,4	S18	8,8	9,0	9,2	9,0	9,4	9,5
S4	7,5	7,9	8,0	8,5	8,8	9,0	S19	5,3	6,5	6,8	6,5	4,8	4,5
S5	8,3	9,5	9,6	9,6	9,5	9,4	S20	9,3	8,9	9,6	9,6	9,4	8,5
S6	9,4	9,4	9,4	9,1	9,3	9,5	S21	6,1	6,8	7,0	7,5	7,6	7,8
S7	3,4	4,0	5,6	6,8	5,4	5,0	S22	8,1	8,1	8,5	8,5	8,3	9,5
S8	5,9	6,4	7,4	5,0	3,8	3,4	S23	5,4	6,1	6,7	7,1	6,7	6,8
S9	9,3	9,5	9,6	9,5	9,8	9,7	S24	8,6	8,2	8,4	8,5	8,8	9,5
S10	3,4	5,1	6,4	5,4	4,4	2,9	S25	5,3	7,2	8,3	5,6	5,3	4,3
S11	3,5	4,3	6,2	4,6	4,1	3,5	S26	4,6	5,4	5,5	5,7	5,8	6,3
S12	7,9	7,6	8,5	8,3	8,4	8,5	S27	5,9	6,3	6,4	6,3	6,8	7,4
S13	8,6	8,3	8,2	8,1	7,6	7,4	S28	6,9	8,3	7,6	7,1	7,4	5,8
S14	9,8	9,6	9,9	9,5	9,8	9,7	S29	7,8	8,2	8,5	8,8	8,7	9,0
S15	7,4	8,2	8,6	8,8	9,0	9,2	S30	5,7	6,0	6,2	6,8	7,0	7,2



Hình 3. Giao diện đồ họa người dùng và hình ảnh của số liệu HS (S1).



Hình 4. Giao diện đồ họa người dùng và hình ảnh của số liệu HS (S3).

Tiếp theo là phần mô tả cách tính từng bước cho số liệu HS S3 dựa trên sự kết hợp giữa phương pháp gần đúng Taylor và mô hình xám GM(2,1). Số liệu thô của HS S3  $x^{(0)} = (4,3; 6,6; 8,4; 6,9; 5,9; 5,4)$ , áp dụng công thức (14) sẽ tính được  $x^{(1)} = (4,3; 10,9; 19,3; 26,2; 32,1; 37,5)$  và công thức (26) tính được  $z^{(1)} = (7,6; 15,1,8; 22,8; 29,2; 34,8)$ . Sau đó sử dụng công thức (23) sẽ tính được các tham số  $a_1$ ,  $a_2$  và  $b$  ( $a_1 = 0,0365$ ,  $a_2 = 0,1302$  và  $b = 3,3121$ ). Sau khi tính được  $a_1$ ,  $a_2$  và  $b$  thì thay vào công thức (18, 19, 20) sẽ tính được các giá trị dự báo của mô hình GM(2,1)  $\hat{x}^{(0)} = (4,3; 3,3; 5,4; 6,8; 7,3; 6,8; 5,5)$ . Từ kết quả  $\hat{x}^{(0)}$  có thể thấy được KQHT của HS S3 dự báo cho học kì tiếp theo là 5,5. Sử dụng

kết quả dự báo so sánh với số liệu thực tế để phân tích sai số cho mô hình xám GM(2,1) dựa theo công thức (41), kết quả sai số MAPE = 22,55%.

Tuy nhiên, khi sử dụng mô hình kết hợp T-GM(1,1) với các hệ số  $K=100$ ,  $h=500$  và  $H=20$ . Kết quả tính được các giá trị dự báo của mô hình T-GM(2,1)  $\hat{x}^{(0)} = (4,3; 6,7; 8,1; 7,1; 6,1; 5,2; 4,4)$ . Từ kết quả  $\hat{x}^{(0)}$  có thể thấy được KQHT của HS S3 dự báo cho học kì tiếp theo là 4,4 và kết quả sai số MAPE = 2,43%. Kết quả trên có thể thấy được trên giao diện đồ họa người dùng khi sử dụng hộp công cụ MATLAB để tính toán (Hình 4). Trên giao diện đồ họa này có thể thấy sai số  $Q$  của mô hình T-GM(2,1) được điều chỉnh giảm dần đến mức tối thiểu.

Bảng 2. Kết quả dự báo và sai số của hai mô hình T-GM(1,1) và T-GM(2,1)

Mã HS	KQHT dự báo	Xếp hạng theo KQHT	Mô hình dự báo	Sai số MAPE (%)
S1	7,29	18	T-GM(1,1)	0,23
S2	9,70	3	T-GM(1,1)	0,01
S4	9,38	9	T-GM(1,1)	0,74
S5	9,43	7	T-GM(1,1)	0,49
S6	9,37	10	T-GM(1,1)	1,04
S9	9,80	1	T-GM(1,1)	0,55
S12	8,77	14	T-GM(1,1)	1,85
S13	7,23	19	T-GM(1,1)	0,97
S14	9,73	2	T-GM(1,1)	1,03
S15	9,50	6	T-GM(1,1)	0,52
S17	8,99	12	T-GM(1,1)	1,28
S18	9,59	5	T-GM(1,1)	0,80
S20	8,91	13	T-GM(1,1)	3,67
S21	8,15	15	T-GM(1,1)	0,79
S22	9,40	8	T-GM(1,1)	2,35
S23	7,10	20	T-GM(1,1)	2,90
S24	9,63	4	T-GM(1,1)	1,32
S26	6,41	21	T-GM(1,1)	1,17
S27	7,48	17	T-GM(1,1)	2,17
S28	5,83	22	T-GM(1,1)	3,39
S29	9,19	11	T-GM(1,1)	0,79
S30	7,65	16	T-GM(1,1)	1,12
S3	4,42	23	T-GM(2,1)	2,43
S7	4,00	24	T-GM(2,1)	5,40
S8	1,97	30	T-GM(2,1)	5,11
S10	2,31	29	T-GM(2,1)	2,76
S11	2,50	28	T-GM(2,1)	5,15
S16	3,86	25	T-GM(2,1)	5,22
S19	3,81	26	T-GM(2,1)	3,31
S25	3,16	27	T-GM(2,1)	5,26

### 3.3. Thảo luận

Dựa theo kết quả ở Bảng 2 cho thấy khi sử dụng hai mô hình T-GM(1,1) và T-GM(2,1) để dự báo kết quả sai số (MAPE) của 30 số liệu đều nhỏ hơn 5,50% điều này chứng tỏ dữ liệu đạt yêu cầu tương đối tốt khi sử dụng hai mô hình dự báo T-GM(1,1) và T-GM(2,1). Từ kết quả này cũng cho thấy cụ thể KQHT môn Sinh học của 30 HS được dự báo cho học kì tiếp theo. Kết quả xếp hạng theo KQHT cho thấy HS S9 có KQHT cao nhất và HS S8 có KQHT thấp nhất. Kết quả này không chỉ là tài liệu tham khảo cho các giáo viên mà còn cung cấp thông tin rất quan trọng cho các nhà quản lí giáo dục để họ chủ động phân loại HS, sắp xếp lớp học hợp lí, đồng thời tuyển chọn HS có thành tích học tập ổn định để bồi dưỡng HS giỏi. Hộp công cụ MATLAB trong nghiên cứu này cũng cho thấy là rất hữu dụng và tiện ích để tính toán và hiển thị các kết quả, hình ảnh một cách trực quan sinh động. Nó giúp cho việc tính toán trở nên nhanh chóng, chính xác và dễ dàng hơn.

Dự báo kết quả học tập của học sinh không chỉ tạo cơ sở khoa học cho việc xây dựng kế hoạch giảng dạy và học tập mà còn cho phép xem xét, đánh giá khả năng học tập của các học sinh trong tương lai. Có thể nói dự báo tốt sẽ cung cấp thông tin rất quan trọng cho giáo viên, học sinh và các nhà quản lí giáo dục để xây dựng chiến lược phát triển giáo dục theo yêu cầu của tương lai. Tóm lại dự báo kết quả học tập của học sinh là một việc làm cần thiết, phương pháp dự báo trong bài viết này giúp cho các nhà giáo dục không phải tốn quá nhiều công nhất nhất là trong điều kiện thông tin không đầy đủ và dữ liệu không đủ lớn để thực hiện các phương pháp thống kê truyền thống.

## 4. Kết luận

Từ kết quả nghiên cứu và thảo luận ở trên cho thấy các tham số của hai mô hình T-GM(1,1) và T-GM(2,1) có thể điều chỉnh cho đến khi đạt giá trị tối ưu và làm cho sai số dự báo giảm đến mức tối thiểu. Kết quả dự báo trong nghiên cứu này có độ chính xác tương đối tốt, kết quả này sẽ cung cấp thông tin rất quan trọng cho GV và cán bộ quản lí giáo dục để giúp cho họ tuyển chọn HS có quá trình học tập ổn định để bồi dưỡng HS giỏi, đồng thời cải thiện KQHT đối với HS có quá trình học tập không ổn định nhằm đáp ứng các yêu cầu và mục tiêu của giáo dục.

Nghiên cứu này cho thấy đã thiết kế thành công một hộp công cụ MATLAB cho hai mô hình dự báo T-GM(1,1) và T-GM(2,1). Hai mô hình T-GM(1,1) và T-GM(2,1) không chỉ sử dụng để dự báo phát triển giáo dục mà còn có thể được sử dụng trong các lĩnh vực khác như y học, kinh tế...

Tóm lại, hai mô hình dự báo này thực sự hữu ích để dự báo cho các hệ thống không chắc chắn khi số lượng dữ liệu là không đủ lớn để sử dụng các phương pháp phân tích thống kê truyền thống và hộp công cụ MATLAB không chỉ giúp cho xử lí dữ liệu một cách nhanh chóng, chính xác, mà còn hiển thị kết quả và hình ảnh rõ ràng trên giao diện đồ họa người dùng.

### Tài liệu tham khảo

- [1] J.L. Deng, Introduction to grey system theory, The Journal of grey system 1 (1989) 1.
- [2] G.D. Li, D. Yamaguchi, K. Mizutani, M. Nagai, New proposal and accuracy evaluation of grey prediction GM, IEICE Transactions on

- Fundamentals of Electronics Communications and Computer Sciences E Series A 90 (2007) 1188.
- [3] G.D. Li, D. Yamaguchi, M. Nagai, S. Masuda, The prediction of asphalt pavement permanent deformation by T-GM (1,2) dynamic model, *International Journal of Systems Science*, 39 (2008) 959.
- [4] C.P. Zhang, Q.Q. Zhou, J. Nie, The Prediction of China CO<sub>2</sub> Emission in 2015, *International Journal of Energy Science* 2 (2012) 47.
- [5] G.D. Li, S. Masuda, M. Nagai, Predictor design using an improved grey model in control systems, *International Journal of Computer Integrated Manufacturing* 1 (2014) 1.
- [6] H.W. Chen, N.B. Chang, Prediction analysis of solid waste generation based on grey fuzzy dynamic modeling, *Resources, conservation and Recycling* 29 (2000) 1.
- [7] S.R. Hui, F. Yang, Z.Z. Li, Q. Liu, J.G. Dong, Application of Grey System Theory to Forecast The Growth of Larch, *International Journal of Information and Systems Sciences* 5 (2009) 522.
- [8] L.J. Liang, L.F. Liu, Y. Li, The Prediction of Shanghai Service Outsourcing Talents Demand Based on Grey Model, *International Journal of Business and Social Science*, 5 (2014) 64.
- [9] Z.X. Liu, B. Wang, K. Xu, H.J. Li, Q.D. Feng, Analysis of China's Water Shortage Model and Relevant Strategies in the Next 20 Years, *Journal of Medical and Bioengineering* 3 (2014) 267.
- [10] G.D. Li, D. Yamaguchi, M. Nagai, Application of improved grey prediction model to short term load forecasting, *Proceedings of International Conference on Electrical Engineering* (2006).
- [11] T.L. Tien, The deterministic grey dynamic model with convolution integral DGDMC(1,n), *Applied Mathematical Modelling* 33 (2009) 3498.
- [12] C.I. Chen, Application of the novel nonlinear grey Bernoulli model for forecasting unemployment rate, *Chaos, Solitons and Fractals* 37 (2008) 278.
- [13] T.L. Tien, A new grey prediction model FGM(1,1), *Mathematical and Computer Modelling* 49 (2009) 1416.
- [14] G.D. Li, S. Masuda, M. Nagai, An Optimal Prediction Model using Taylor Approximation Method, *Journal of Grey System* 11 (2011) 173.
- [15] W. Li, H. Xie, Geometrical Variable Weights Buffer GM(1,1) Model and Its Application in Forecasting of China's Energy Consumption, *Journal of Applied Mathematics* 2014 (2014) 1.
- [16] T.W. Sheu, P.H. Nguyen, P.T. Nguyen, D.H. Pham, C.P. Tsai, M. Nagai, Using the Combination of GM(1,1) and Taylor Approximation Method to Predict the Academic Achievement of Student, *SOP Transactions on Applied Mathematics* 1 (2014) 55.
- [17] T.W. Sheu, P.H. Nguyen, P.T. Nguyen, D.H. Pham, C.P. Tsai, M. Nagai, Using GM(2,1) and T-GM(2,1) to Predict the Number of Students for Admission, *Journal of Information and Computational Science* 11(2014) 6085.
- [18] M. Nagai, D. Yamaguchi, *Grey Theory and Engineering Application Method*. Tokyo: Kyoritsu Publisher, 2004.
- [19] Z.J. Guo, X.Q. Song, J. Ye, A Verhulst model on time series error corrected for port throughput forecasting, *Journal of the Eastern Asia society for Transportation studies* 6 (2005) 881.
- [20] L.D. Qu, D.X. He, R.M. Jia, Optimized Grey Model Based on Cuckoo Search Algorithm and Its Prediction Application, *Journal of Information and Computational Science* 11 (2014) 1419.
- [21] C.N. Wang, V.T. Phan, An improvement the accuracy of grey forecasting model for cargo throughput in international commercial ports of Kaohsiung, *International Journal of Business and Economics Research* 3 (2014) 1.
- [22] K.L. Wen, T.C. Chang, The research and development of completed GM(1,1) model toolbox using Matlab. *International Journal of Computational Cognition* 3 (2005) 42.
- [23] T.W. Sheu, P.H. Nguyen, P.T. Nguyen, D.H. Pham, A Matlab Toolbox for AHP and LGRA-AHP to Analyze and Evaluate Factors in Making the Decision, *International Journal of Kansei Information* 4 (2013) 149.
- [24] T.W. Sheu, P.H. Nguyen, P.T. Nguyen, D.H. Pham, C.P. Tsai, M. Nagai, A MATLAB Toolbox for Misconceptions Analysis Based on S-P Chart, Grey Relational Analysis and ROC, *Transactions on Machine Learning and Artificial Intelligence* 2 (2014) 72.

## Predicting the Student Learning Outcomes Based on the Combination of Taylor Approximation Method and Grey Models

Nguyễn Phước Hải<sup>1</sup>, Tian-Wei Sheu<sup>2</sup>, Masatake Nagai<sup>2</sup>

<sup>1</sup> *Kien Giang Teacher Training College,*

*No. 449, Nguyen Chi Thanh street, Rach Gia city, Kien Giang province, Vietnam*

<sup>2</sup> *Graduate Institute of Educational Information and Measurement,*

*National Taichung University of Education, Taiwan,*

*No. 140, Minsheng Rd., West Dist., Taichung City 40306, Taiwan (R.O.C.)*

**Abstract:** The purpose of this study is to predict the student learning outcomes based on the combination of Taylor approximation method with two grey models GM(1,1) and GM(2,1). Two combined models T-GM(1,1) and T-GM(2,1) can obtain the most optimal predicted values by multi-times approximate calculation to improve the predicted accuracy of two grey models. In addition, researchers have used the MATLAB software to design a MATLAB toolbox for two combined models. The results of this study will provide important information for teachers and education managers to help them select students having the stable learning process to foster good students, improve learning outcomes for students having the unstable learning process to meet the requirements and objectives of education.

**Keywords:** Learning outcomes, Taylor approximation method, grey models, MATLAB toolbox, learning process.