

# On unified framework for continuous-time grey models: An integral matching perspective

Baolei Wei, Naiming Xie\*

College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

## ARTICLE INFO

### Article history:

Received 11 May 2021

Revised 20 August 2021

Accepted 5 September 2021

Available online 15 September 2021

### Keywords:

Grey system model

Cumulative sum operator

Integral matching

Water supply forecast

## ABSTRACT

Since most of the research about grey forecasting models is focused on developing novel models and improving accuracy, relatively limited attention has been paid to the modelling mechanism and relationships among diverse kinds of models. This paper aims to unify and reconstruct continuous-time grey models, highlighting the differences and similarities among different models. First, the unified form of grey forecasting models is proposed and simplified into a reduced-order ordinary differential equation. Then, the integral matching that consists of integral operator and least squares, is proposed to estimate the structural parameter and initial value simultaneously. The cumulative sum operator, an essential element in grey modelling, proves to be the discrete approximation of the integral operator. Next, grey models are reconstructed by the integral matching-based ordinary differential equations. Finally, the existing grey models are compared with the reconstructed models through extensive simulations, and a real-world example shows how to apply and further verify the reconstructed model.

© 2021 Elsevier Inc. All rights reserved.

## 1. Introduction

Over the past four decades since the seminal work on grey system theory [1], grey system models have been widely used to various fields from natural science through social science. Grey forecasting models which are pioneered in 1984 [2], are constructed based on the prior assumption that the accumulation and release in many generalized energy systems conform to an exponential law [3]. Grey forecasting models utilize the Cusum (cumulative sum) operator to mine the exponential characteristic hidden in the original time series and then the continuous-time dynamics (ordinary differential equations) are employed to fit the Cusum series. Generally, grey forecasting models, also refereed as to Grey Models in some literature, are abbreviated as  $GM(\varphi, \psi)$ , where  $\varphi$  denotes the order of derivative and  $\psi$  denotes the number of variables. Research has shown consistently that grey forecasting models are promising in solving a class of time series prediction problems, especially in the case of small-sample data sets [4].

In the following, we review grey forecasting models from three viewpoints: the Cusum operator, the basic model and the extended ones.

\* Corresponding author.

E-mail addresses: [weibl@nuaa.edu.cn](mailto:weibl@nuaa.edu.cn) (B. Wei), [xienaiming@nuaa.edu.cn](mailto:xienaiming@nuaa.edu.cn) (N. Xie).

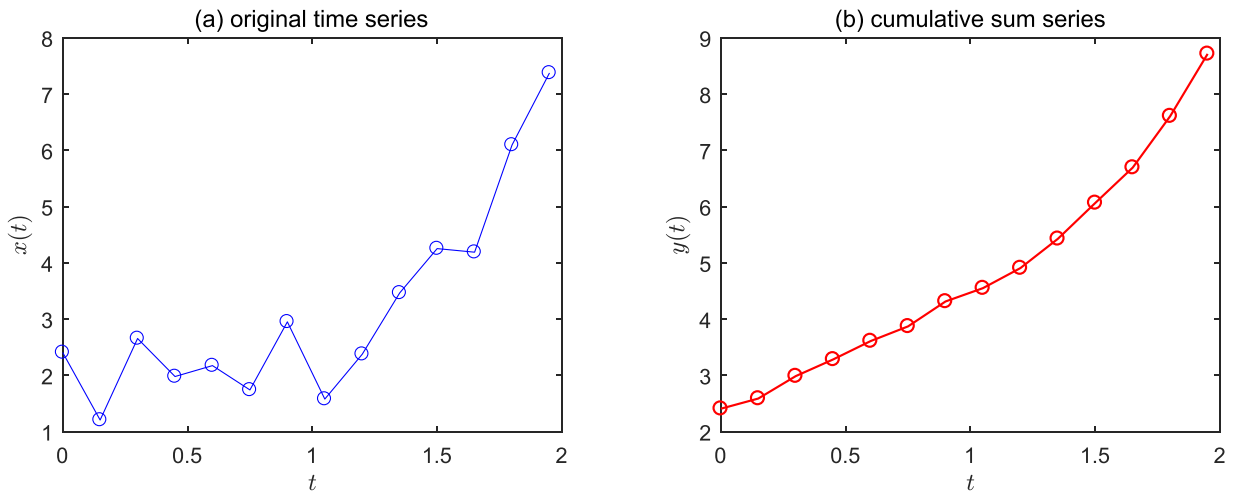


Fig. 1. Plots of the original time series and the corresponding cumulative sum series.

### 1.1. The cumulative sum operator and its inverse

**Definition 1** (See [5]). For the original time series  $X(t) = \{x(t_1), x(t_2), \dots, x(t_n)\}$ , the Cusum series is defined as  $Y(t) = \{y(t_1), y(t_2), \dots, y(t_n)\}$ , where

$$y(t_k) = \sum_{i=1}^k h_i x(t_i), \quad h_k = \begin{cases} 1, & k = 1 \\ t_k - t_{k-1}, & k \geq 2 \end{cases}.$$

Correspondingly, the inverse Cusum operator is defined as

$$x(t_k) = \begin{cases} \frac{1}{h_1} y(t_1), & k = 1 \\ \frac{1}{h_k} [y(t_k) - y(t_{k-1})], & k \geq 2 \end{cases}.$$

Note that the first time interval  $h_1$  is set to 1 according to the empty product rule in pure mathematics [6]. The Cusum operator associated with its inverse is an omni-directional formula for not only the equally spaced time series but the irregularly spaced ones [5,7,8].

Definition 1 shows that the original time series can be obtained by restoring the Cusum series. Inspired by this idea, grey forecasting models were first proposed by fitting and forecasting the Cusum series and then restoring the results [2]. In a broad sense, the Cusum operator can be viewed as a data preprocessing technique to non-parametrically mine the pattern hidden in the original time series. There are some other advantages of Cusum operator, especially from a data visualization perspective. For example, it is difficult to identify the hidden pattern from the line graph of the original time series (see Fig. 1(a)), while the quasi-exponential characteristic is obvious in the cumulative sum series (see Fig. 1(b)).

In fact, grey forecasting models are not the only ones that use the Cusum operator as a pattern recognition tool. The similar idea was also employed in other disciplines, such as ecology [9] and statistical process control [10].

### 1.2. The basic grey forecasting model

The GM(1,1) model, employing the single-variable and first-order ordinary differential equation, is the most basic and popular grey forecasting model which has been substantiated in many fields. It consists of the continuous-time differential equation

$$\frac{d}{dt}y(t) = ay(t) + b, \quad t \geq t_1 \quad (1)$$

and the discrete-time difference equation

$$x(t_k) = a[\lambda y(t_{k-1}) + (1 - \lambda)y(t_k)] + b \quad (2)$$

where  $k = 2, 3, \dots, n$ ,  $n \geq 4$ , and  $\lambda \in [0, 1]$  is a hyperparameter (also referred to as a background coefficient) whose value should be set ahead of the following parameter estimation. By using the least squares, the structural parameters are estimated from Eq. (2) as

$$\begin{bmatrix} \hat{a} & \hat{b} \end{bmatrix}^T = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{y} \quad (3)$$

where

$$D = \begin{bmatrix} \lambda y(t_1) + (1 - \lambda)y(t_2) & 1 \\ \lambda y(t_2) + (1 - \lambda)y(t_3) & 1 \\ \vdots & \vdots \\ \lambda y(t_{n-1}) + (1 - \lambda)y(t_n) & 1 \end{bmatrix}, \quad y = \begin{bmatrix} x(t_2) \\ x(t_3) \\ \vdots \\ x(t_n) \end{bmatrix}.$$

By setting the initial value parameter to

$$y(t_1) = \eta, \quad (4)$$

the time response function is calculated as

$$\hat{y}(t) = \left( \eta + \frac{\hat{b}}{\hat{a}} \right) e^{\hat{a}(t-t_1)} - \frac{\hat{b}}{\hat{a}}, \quad t \geq t_1, \quad (5)$$

then by using the inverse Cusum operator, the forecasting results corresponding to the original time series are obtained as  $\hat{x}(t_1) = \eta$  and  $\hat{x}(t_k) = \frac{1}{h_k} [\hat{y}(t_k) - \hat{y}(t_{k-1})]$ ,  $k = 2, 3, \dots, n + r$ , where  $r$  is the forecasting horizon.

On the basis of the above modelling process, the existing studies for improving GM(1,1) model can be divided into the following types:

- (i) Eq. (2) is the numerical discretization-based approximation of Eq. (1) by using the weighted trapezoid rule which yields the classical trapezoid formula with the background coefficient equal to 0.5. Researchers have been making efforts to improve accuracy by optimizing the background coefficient, such as using a differential evolution algorithm to search the optimal background coefficient [11]. On the other hand, by using the trapezoid formula in each closed interval  $[t_{k-1}, t_k]$ , the discrete-time Eq. (2) is modified to be the following form

$$x(t_k) = a[\lambda_k y(t_{k-1}) + (1 - \lambda_k)y(t_k)] + b$$

where  $\lambda_k \in [0, 1]$ ,  $k = 2, 3, \dots, n$ . In each closed interval  $[t_{k-1}, t_k]$ , the background coefficient  $\lambda_k$  is adaptively computed by using the triangular membership function rule [12–14] or the exponential function rule [15].

- (ii) The initial value was set to  $\hat{y}(t_1) = x(t_1)$  in the classical research [3]. Therefore, based on the intuition that new observations are likely to be more informative than historical ones, an alternative strategy for the initial value selection is  $\hat{y}(t_n) = y(t_n)$  [16]. On the other hand, the least-square strategy [17] was proposed to further reduce the fitting error:

$$\hat{\eta} = \arg \min_{\eta} \sum_{k=1}^n (y(t_k) - \hat{y}(t_k))^2$$

where  $\hat{y}(t_k)$  is calculated according to the time response Eq. (5) which has estimated structural parameters but unknown initial value.

- (iii) Combining the background coefficient optimization and initial value selection together is a straightforward approach to improve modelling accuracy. For example, a hybrid method combined an explicit formula for background coefficient calculation and the least-square strategy for initial value selection [18].

In addition, another research focus of GM(1,1) model is the property analysis, such as the necessary and sufficient condition for modelling [19], the error bound estimation among the variants of GM(1,1) model [20], the influence of data transformation on modelling accuracy [7,21] and the effect of sample size on modelling performance [22,23].

### 1.3. The extended grey forecasting models

Since the procedures of GM(1,1) provide a clear paradigm for grey modelling, Guo and Guo [24] probed into the mathematical principles for model extensions but this is short of details. A rough classification of the existing outputs shows that the extended models are mainly focused on the following aspects: (i) extending the linear model into the nonlinear ones; (ii) extending the signal-variable model into the multi-variable and multi-output ones.

First, the extended single-variable models have a similar modelling procedures with GM(1,1) model, i.e., modelling the Cusum series with ordinary differential equations. The main differences lie in the continuous-time and discrete-time equations. Table 1 shows the linear extensions have similar expressions, excluding the forcing terms in differential equations.

Table 1 shows that the extensions have an inclusion relation, i.e., the former is a special case of the latter in the linear and nonlinear cases, respectively. In particular, both GPM(1,1,N) and KRNGM(1,1) are extensions from a basis expansion perspective; the former uses the polynomial basis function, while the latter utilizes the kernel function. The Grey Verhulst model, aimed at fitting the inverted U-shaped series, is the first nonlinear extension, and then the power coefficient is generalized from the fixed integer 2 to a real number, in order to improve the flexibility in NBGM(1,1) model [32].

Second, the multi-variable extensions corresponding to the single-variable models were proposed successively. The first multi-variable extension GM(1,N) plays an important role in establishing the multi-variable models, even though it is an incomplete extension of GM(1,1) [3]; then on this basis, Tien [33] came to the complete extension by introducing a control

**Table 1**

The coupled equations of the extended single-variable linear and nonlinear grey forecasting models.

Type	Model	Differential equation	Difference equation	Ref
Linear	GM(1,1)	$\frac{d}{dt}y(t) = ay(t) + b$	$x(k) = a \frac{y(k-1)+y(k)}{2} + b$	[3]
	NGM(1,1,k)	$\frac{d}{dt}y(t) = ay(t) + bt$	$x(k) = a \frac{y(k-1)+y(k)}{2} + bk$	[25]
	NGM(1,1,k, c)	$\frac{d}{dt}y(t) = ay(t) + b_1t + b_0$	$x(k) = a \frac{y(k-1)+y(k)}{2} + b_1k + b_0$	[26]
	GM(1,1,t $^\alpha$ )	$\frac{d}{dt}y(t) = ay(t) + b_1t^\alpha + b_0$	$x(k) = a \frac{y(k-1)+y(k)}{2} + b_1k^\alpha + b_0$	[27]
	GPM(1,1,N)	$\frac{d}{dt}y(t) = ay(t) + \sum_{j=0}^N b_j t^j$	$x(k) = a \frac{y(k-1)+y(k)}{2} + \sum_{j=0}^N b_j \frac{k^{j+1} - (k-1)^{j+1}}{j+1}$	[28]
	KRNGM(1,1)	$\frac{d}{dt}y(t) = ay(t) + \sum_{j=0}^N b_j \phi_j(t)$	$x(k) = a \frac{y(k-1)+y(k)}{2} + \sum_{j=0}^N b_j \frac{\phi_j(k) + \phi_j(k-1)}{2}$	[29]
Nonlinear	Verhulst	$\frac{d}{dt}y(t) = ay(t) + b[y(t)]^2$	$x(k) = a \frac{y(k-1)+y(k)}{2} + b \left( \frac{y(k-1)+y(k)}{2} \right)^2$	[30]
	NBGM(1,1)	$\frac{d}{dt}y(t) = ay(t) + b[y(t)]^\gamma$	$x(k) = a \frac{y(k-1)+y(k)}{2} + b \left( \frac{y(k-1)+y(k)}{2} \right)^\gamma$	[31]

Note that the time interval in these models is set to unit, i.e.,  $t_k = k$ ,  $k = 1, 2, \dots, n$ .

parameter. Subsequently, researchers proposed some other multi-variable models including the NGM(1,N) [34] which is the extension of NGM(1,1,k, c), the KGM(1,N) [35] which is the extension of KRNGM(1,1), and the NBGM(1,N) [36,37] which is the extension of NBGM(1,1) model.

Third, further extension of multi-variable models leads to the multi-output ones, but the literature is generally sparse on this up to now. To the best of our knowledge, these mainly include the linear MGM(1,N) model [38] associated with its improvement coupled with self-memory [39], and the nonlinear MNBGM(1,N) model coupled with self-memory [40], as well as the nonlinear grey Lotka–Volterra model [41,42].

Finally, in a similar manner to the research route of the GM(1,1) model, there have been a variety of optimization studies on the aforementioned three categories of extension models, such as the background coefficient and/or initial value optimization for (i) the single-variable cases such as the GPM(1,1,N) model [43] and the NBGM(1,1) model [44] and (ii) the multi-variable cases such as the NGM(1,N) model [45].

#### 1.4. Motivation

Having reviewed a wide range of grey forecasting models in the previous introductory section, the next sections will address, in detail, certain challenging issues. The main objectives of this paper are as follows:

- There exist a variety of grey forecasting models with different model representations (see Table 1 and the multi-variable and multi-output models that follow), making it difficult for researchers to perform property analysis; and for practitioners to select an appropriate model for a given practical problem.
- The Cusum operator is a fundamental element in the grey forecasting model but, until now, the mechanism has not been explained clearly, especially from a mathematical perspective.
- Forecasting the future values of the original time series is the main goal of a grey model, so is it possible to model the original time series directly but achieve the same performance without explicit use of Cusum operator? In this way, the modelling procedures will be simplified because it avoids the inverse Cusum operator when calculating the forecasts of the original time series, thus making the modelling results easier to explain.

The principal contributions of the present research are as follows:

- The aforementioned linear grey forecasting models, together with their multi-variable and multi-output extensions, are unified by a matrix differential equation, after extending the Cusum operator for single-variable time series to one that is suitable for vector time series.
- By introducing an integral operator, the Cusum operator proves to be the Euler's formula-based numerical discretization of this operator; and, subsequently, the mechanism of the Cusum operator is further explained within the parameter estimation process.
- The unified grey forecasting model for Cusum series proves to be equivalent to a reduced-order ordinary differential equation for the original series. This provides the basis for reconstructing grey forecasting models using an integral matching-based ordinary differential equation framework, in which the differential equation is first converted to an integral equation; and then the structural parameter and initial value are estimated simultaneously.

The remaining parts of the paper are organized as follows: in Section 2, we extend the Cusum operator, propose a unified representation, reconstruct the modelling process using matrix analysis, and discuss some special cases of the unified model; in Section 3, we introduce the integral operator to explain the Cusum mechanism, obtain the equivalent reduced-order differential equation, and present integral matching as a method for simultaneously estimating the structural parameter and initial value; in Section 4, we analyze the relationship between the grey model and the integral matching-based one, then

reconstruct the modelling procedures of the grey model; Section 5 provides large-scale simulations that compare the two approaches from both parameter estimation and forecasting accuracy viewpoints; Section 6 provides a real-world example to show how to apply our method; while Section 7 concludes the work and discusses the future directions of research and development.

## 2. Unified representation of grey forecasting models

In this section we present a universal framework for the existing linear grey forecasting models, including structural parameter estimation and initial value selection strategies.

For a state vector  $\mathbf{x}(t) \in \mathbb{R}^d$ , the observational data sampled at time instants  $t_1, t_2, \dots, t_n$  are arranged into the following matrix form:

$$\begin{bmatrix} \mathbf{x}^T(t_1) \\ \mathbf{x}^T(t_2) \\ \vdots \\ \mathbf{x}^T(t_n) \end{bmatrix} = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \cdots & x_d(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots & x_d(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_n) & x_2(t_n) & \cdots & x_d(t_n) \end{bmatrix}$$

and subsequently the corresponding cumulative sum form is expressed as

$$\begin{bmatrix} \mathbf{y}^T(t_1) \\ \mathbf{y}^T(t_2) \\ \vdots \\ \mathbf{y}^T(t_n) \end{bmatrix} = \begin{bmatrix} y_1(t_1) & y_2(t_1) & \cdots & y_d(t_1) \\ y_1(t_2) & y_2(t_2) & \cdots & y_d(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ y_1(t_n) & y_2(t_n) & \cdots & y_d(t_n) \end{bmatrix}$$

where the Cusum column vector is calculated according to the formula  $\mathbf{y}(t_k) = \sum_{i=1}^k h_i \mathbf{x}(t_i)$  which is the multi-variable extension of the single-variable Cusum operator in Definition 1.

For the multi-variable Cusum series, consider the grey forecasting models involving the ordinary differential equation

$$\frac{d}{dt} \mathbf{y}(t) = \mathbf{A} \mathbf{y}(t) + \mathbf{B} \mathbf{u}(t) + \mathbf{c}, \quad t \geq t_1 \quad (6)$$

where  $\mathbf{y}(t) \in \mathbb{R}^d$  is the Cusum vector,  $\mathbf{u}(t) \in \mathbb{R}^p$  is a known vector (that is independent of  $\mathbf{y}(t)$ ),  $\mathbf{A} \in \mathbb{R}^{d \times d}$ ,  $\mathbf{B} \in \mathbb{R}^{d \times p}$  and  $\mathbf{c} \in \mathbb{R}^d$  are unknown structural parameters.

First, numerical discretization-based gradient matching is employed to estimate the structural parameters in Eq. (6). This method can be divided into the following two steps: in the first step, using the trapezoid formula gives the discrete-time equation

$$\frac{\mathbf{y}(t_k) - \mathbf{y}(t_{k-1})}{t_k - t_{k-1}} = \mathbf{A} \frac{\mathbf{y}(t_{k-1}) + \mathbf{y}(t_k)}{2} + \mathbf{B} \frac{\mathbf{u}(t_{k-1}) + \mathbf{u}(t_k)}{2} + \mathbf{c} + \mathcal{O}(h^2) \quad (7)$$

where  $h = \max_{k=2}^n \{h_k\}$ . Realistically, often only  $\mathbf{x}(t)$  is available and contaminated with noise. Thus,  $\mathbf{y}(t)$  is contaminated with noise and Eq. (7) does not hold exactly. By substituting  $\mathbf{x}(t_k)$  for  $\frac{\mathbf{y}(t_k) - \mathbf{y}(t_{k-1})}{t_k - t_{k-1}}$ , the expression (7) can be rewritten as

$$\mathbf{x}(t_k) = \mathbf{A} \frac{\mathbf{y}(t_{k-1}) + \mathbf{y}(t_k)}{2} + \mathbf{B} \frac{\mathbf{u}(t_{k-1}) + \mathbf{u}(t_k)}{2} + \mathbf{c} + \boldsymbol{\varepsilon}(t_k) \quad (8)$$

where  $\boldsymbol{\varepsilon}(t_k)$  is the sum of discretization error and noise error. By substituting  $k = 2, 3, \dots, n$  into Eq. (8) and arranging the resulting  $n - 1$  equations into a matrix form, one has

$$\mathbf{X} = \boldsymbol{\Theta}(\mathbf{y}, \mathbf{u}) \boldsymbol{\Xi} + \boldsymbol{\Gamma} \quad (9)$$

where

$$\boldsymbol{\Xi} = \begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \\ \mathbf{c}^T \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}^T(t_2) \\ \mathbf{x}^T(t_3) \\ \vdots \\ \mathbf{x}^T(t_n) \end{bmatrix}, \quad \boldsymbol{\Theta}(\mathbf{y}, \mathbf{u}) = \begin{bmatrix} \frac{\mathbf{y}^T(t_1) + \mathbf{y}^T(t_2)}{2} & \frac{\mathbf{u}^T(t_1) + \mathbf{u}^T(t_2)}{2} & 1 \\ \frac{\mathbf{y}^T(t_2) + \mathbf{y}^T(t_3)}{2} & \frac{\mathbf{u}^T(t_2) + \mathbf{u}^T(t_3)}{2} & 1 \\ \vdots & \vdots & \vdots \\ \frac{\mathbf{y}^T(t_{n-1}) + \mathbf{y}^T(t_n)}{2} & \frac{\mathbf{u}^T(t_{n-1}) + \mathbf{u}^T(t_n)}{2} & 1 \end{bmatrix}, \quad \boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\varepsilon}^T(t_2) \\ \boldsymbol{\varepsilon}^T(t_3) \\ \vdots \\ \boldsymbol{\varepsilon}^T(t_n) \end{bmatrix}.$$

Then, in a second step, we obtain the numerical discretization-based estimate  $\hat{\boldsymbol{\Xi}}$  of  $\boldsymbol{\Xi}$  by minimizing the least-squares criterion

$$\mathcal{L}(\boldsymbol{\Xi}) = \|\mathbf{X} - \boldsymbol{\Theta}(\mathbf{y}, \mathbf{u}) \boldsymbol{\Xi}\|_F^2 \quad (10)$$

where  $\|\cdot\|_F$  is the Frobenius norm. The condition for this is obtained in the usual manner by partially differentiating the objective function  $\mathcal{L}(\boldsymbol{\Xi})$  with respect to the parameter matrix, in turn, and setting the derivative to zero. This yields the matrix equation

$$\frac{\partial}{\partial \boldsymbol{\Xi}} \mathcal{L}(\boldsymbol{\Xi}) = \frac{\partial}{\partial \boldsymbol{\Xi}} \text{Tr}([\mathbf{X} - \boldsymbol{\Theta}(\mathbf{y}, \mathbf{u}) \boldsymbol{\Xi}]^T [\mathbf{X} - \boldsymbol{\Theta}(\mathbf{y}, \mathbf{u}) \boldsymbol{\Xi}]) = 2\boldsymbol{\Theta}^T(\mathbf{y}, \mathbf{u}) \boldsymbol{\Theta}(\mathbf{y}, \mathbf{u}) \boldsymbol{\Xi} - 2\boldsymbol{\Theta}^T(\mathbf{y}, \mathbf{u}) \mathbf{X} = 0$$

and the subsequent least-square estimate

$$\hat{\Xi} = (\Theta^T(\mathbf{y}, \mathbf{u})\Theta(\mathbf{y}, \mathbf{u}))^{-1}\Theta^T(\mathbf{y}, \mathbf{u})\mathbf{X}. \quad (11)$$

Next, we solve Eq. (6) using the estimated parameters. By employing the variation of parameters method, one has the general solution (also called time response function) to the non-homogeneous Eq. (6) expressed as

$$\tilde{\mathbf{y}}(t) = \exp(\hat{\mathbf{A}}(t - t_1)) \left\{ \boldsymbol{\eta} + \int_{t_1}^t \exp(\hat{\mathbf{A}}(t_1 - s)) [\hat{\mathbf{B}}\mathbf{u}(s) + \hat{\mathbf{c}}] ds \right\} \quad (12)$$

where  $\tilde{\mathbf{y}}(t_1) = \boldsymbol{\eta} \in \mathbb{R}^d$  is the unknown initial vector. In order to forecast the evolutions, the required initial vector  $\boldsymbol{\eta}$  is always determined by using one of the following strategies:

(i) the fixed first point strategy assuming

$$\tilde{\mathbf{y}}(t) = \mathbf{y}(t_1) \Rightarrow \boldsymbol{\eta} = \mathbf{y}(t_1),$$

which leads to the time response function

$$\hat{\mathbf{y}}(t) = \exp(\hat{\mathbf{A}}(t - t_1)) \left\{ \mathbf{y}(t_1) + \int_{t_1}^t \exp(\hat{\mathbf{A}}(t_1 - s)) [\hat{\mathbf{B}}\mathbf{u}(s) + \hat{\mathbf{c}}] ds \right\}; \quad (13)$$

(ii) the fixed last point strategy assuming

$$\tilde{\mathbf{y}}(t_n) = \mathbf{y}(t_n) \Rightarrow \boldsymbol{\eta} = \exp(\hat{\mathbf{A}}(t_1 - t_n))\mathbf{y}(t_n) - \int_{t_1}^{t_n} \exp(\hat{\mathbf{A}}(t_1 - s)) [\hat{\mathbf{B}}\mathbf{u}(s) + \hat{\mathbf{c}}] ds,$$

which leads to the time response function

$$\hat{\mathbf{y}}(t) = \exp(\hat{\mathbf{A}}(t - t_1)) \left\{ \exp(\hat{\mathbf{A}}(t_1 - t_n))\mathbf{y}(t_n) + \int_{t_n}^t \exp(\hat{\mathbf{A}}(t_1 - s)) [\hat{\mathbf{B}}\mathbf{u}(s) + \hat{\mathbf{c}}] ds \right\}; \quad (14)$$

(iii) the least-squares strategy minimizing the least-square objective function

$$\tilde{\mathbf{y}}(t_1) = \hat{\boldsymbol{\eta}} = \arg \min_{\boldsymbol{\eta}} \left\{ \mathcal{L}(\boldsymbol{\eta}) = \sum_{k=1}^n \|\mathbf{y}(t_k) - \tilde{\mathbf{y}}(t_k)\|_2^2 \right\},$$

which subsequently leads to the time response function

$$\hat{\mathbf{y}}(t) = \exp(\hat{\mathbf{A}}(t - t_1)) \left\{ \hat{\boldsymbol{\eta}} + \int_{t_1}^t \exp(\hat{\mathbf{A}}(t_1 - s)) [\hat{\mathbf{B}}\mathbf{u}(s) + \hat{\mathbf{c}}] ds \right\}. \quad (15)$$

Eqs. (13), (14) and (15) show the most commonly-used strategies to determine the initial vector. Finally, substituting the time points  $\{t_k\}_{k=1}^{n+r}$  into Eq. (13), (14) or (15) yields the fitting and forecasting values of the Cusum series and then, using the inverse Cusum operator, we obtain the fitting and forecasting results of the original multi-variable time series:

$$\hat{\mathbf{x}}(t_k) = \begin{cases} \hat{\mathbf{y}}(t_1), & k = 1 \\ \frac{1}{t_k - t_{k-1}} [\hat{\mathbf{y}}(t_k) - \hat{\mathbf{y}}(t_{k-1})], & k = 2, 3, \dots, n + r \end{cases} \quad (16)$$

Overall, the modelling procedures of grey forecasting models can be summarized easily from the above process. Then, we are able to analyze this unified model from three different scenarios including the single-variable, multi-variable and multi-output ones.

**Remark 1.** Let the dimension of the state vector be  $d = 1$ . If  $\mathbf{u}(t) = [u_i(t)]_{p \times 1}$  consists of the fixed basis functions of time, then the unified model yields multiple families of single-variable grey forecasting models.

(1) For the family of grey polynomial models, the basis functions are  $u_i(t) = t^i$ ,  $i = 1, 2, \dots, p$ . By inserting  $u_i(t)$  back into Eqs. (6) and (8), we obtain the continuous- and discrete-time equations respectively expressed as

$$\frac{d}{dt}y(t) = ay(t) + \sum_{i=1}^p b_i t^i + c$$

and

$$x(t_k) = a \frac{y(t_{k-1}) + y(t_k)}{2} + \sum_{i=1}^p b_i \frac{t_{k-1}^i + t_k^i}{2} + c + \varepsilon(t_k).$$

- (2) For the family of grey Fourier models, the basis functions are  $u_{2l-1}(t) = \sin(2l\pi ft)$  and  $u_{2l}(t) = \cos(2l\pi ft)$ ,  $l = 1, 2, \dots, \frac{p}{2}$ , where  $p$  is even and  $f$  is frequency. By inserting  $u_{2l-1}(t)$  and  $u_{2l}(t)$  back into Eqs. (6) and (8), we obtain the continuous- and discrete-time equations respectively expressed as

$$\frac{d}{dt}y(t) = ay(t) + \sum_{l=1}^{p/2} b_{2l-1} \sin(2l\pi ft) + \sum_{l=1}^{p/2} b_{2l} \cos(2l\pi ft) + c$$

and

$$\begin{aligned} x(t_k) = & a \frac{y(t_{k-1}) + y(t_k)}{2} + \sum_{l=1}^{p/2} b_{2l-1} \frac{\sin(2l\pi ft_{k-1}) + \sin(2l\pi ft_k)}{2} \\ & + \sum_{l=1}^{p/2} b_{2l} \frac{\cos(2l\pi ft_{k-1}) + \cos(2l\pi ft_k)}{2} + c + \varepsilon(t_k). \end{aligned}$$

**Remark 1** shows that this unified model subsumes a number of special families of grey models, although here only two of them are exemplified. Furthermore, each family also subsumes many specific models that have been researched and received specific names. For instance, all the linear models in Table 1 belong to the grey polynomial model family and, additionally, the grey Fourier model family [46] covers the hybrid grey models which uses the Fourier series fitting to modify the residual errors of GM(1,1) model [47–50].

**Remark 2.** Let the dimension of the state vector be  $d = 1$ . If the vector  $\mathbf{u}(t) = [u_l(t)]_{p \times 1}$  consists of other state variables, i.e.,  $u_l(t) = y_l(t)$ ,  $l = 1, 2, \dots, p$ , then the unified model yields the multi-variable grey forecasting model [33,51], whose continuous- and discrete-time equations are

$$\frac{d}{dt}y(t) = ay(t) + \sum_{l=1}^p b_l y_l(t) + c$$

and

$$x(t_k) = a \frac{y(t_{k-1}) + y(t_k)}{2} + \sum_{l=1}^p b_l \frac{y_l(t_{k-1}) + y_l(t_k)}{2} + c + \varepsilon(t_k).$$

If the vector  $\mathbf{u}(t) = [u_l(t)]_{p \times 1}$  consists of a mixture of other state variables and known fixed basis functions of time, then combining the extension principle in Remark 1 gives

$$\frac{d}{dt}y(t) = ay(t) + \sum_{l=1}^{p_1} b_l y_l(t) + \sum_{l=1}^{p_2} b_l u_l(t) + c$$

and

$$x(t_k) = a \frac{y(t_{k-1}) + y(t_k)}{2} + \sum_{l=1}^{p_1} b_l \frac{y_l(t_{k-1}) + y_l(t_k)}{2} + \sum_{l=1}^{p_2} b_l \frac{u_l(t_{k-1}) + u_l(t_k)}{2} + c + \varepsilon(t_k),$$

which shows that this model yields many multi-variable grey forecasting models, such as the aforementioned NGM(1,N) model [34] ( $p_1 = N - 1$ ,  $p_2 = 1$  and  $u_1(t) = t$ ). This model may also lead to many other novel multi-variable grey forecasting models.

**Remark 3.** Let the dimension of the state vector be  $d \geq 2$ . If the vector  $\mathbf{u}(t) = [u_l(t)]_{p \times 1}$  is equal to  $\mathbf{0}$ , i.e.,  $u_l(t) = 0$ ,  $l = 1, 2, \dots, p$ , then the unified model yields the multi-output grey forecasting model [38,52], whose continuous- and discrete-time equations are

$$\begin{cases} \frac{d}{dt}y_1(t) = \sum_{l=1}^d a_{1,l}y_l(t) + c_1 \\ \frac{d}{dt}y_2(t) = \sum_{l=1}^d a_{2,l}y_l(t) + c_2 \\ \vdots \\ \frac{d}{dt}y_d(t) = \sum_{l=1}^d a_{d,l}y_l(t) + c_d \end{cases}$$

and

$$\begin{cases} x_1(t_k) = \sum_{l=1}^d a_{1,l} \frac{y_l(t_{k-1}) + y_l(t_k)}{2} + c_1 + \varepsilon_1(t_k) \\ x_2(t_k) = \sum_{l=1}^d a_{2,l} \frac{y_l(t_{k-1}) + y_l(t_k)}{2} + c_2 + \varepsilon_2(t_k) \\ \vdots \\ x_d(t_k) = \sum_{l=1}^d a_{d,l} \frac{y_l(t_{k-1}) + y_l(t_k)}{2} + c_d + \varepsilon_d(t_k) \end{cases}.$$

Similar to the extensions in [Remarks 1](#) and [2](#), if the vector  $\mathbf{u}(t) = [u_l(t)]_{p \times 1}$  consists of fixed functions of time, then the continuous- and discrete-time equations are expressed as

$$\begin{cases} \frac{d}{dt} y_1(t) = \sum_{l=1}^d a_{1,l} y_l(t) + \sum_{l=1}^p b_{1,l} u_l(t) + c_1 \\ \frac{d}{dt} y_2(t) = \sum_{l=1}^d a_{2,l} y_l(t) + \sum_{l=1}^p b_{2,l} u_l(t) + c_2 \\ \vdots \\ \frac{d}{dt} y_d(t) = \sum_{l=1}^d a_{d,l} y_l(t) + \sum_{l=1}^p b_{d,l} u_l(t) + c_d \end{cases}$$

and

$$\begin{cases} x_1(t_k) = \sum_{l=1}^d a_{1,l} \frac{y_l(t_{k-1}) + y_l(t_k)}{2} + \sum_{l=1}^p b_{1,l} \frac{u_l(t_{k-1}) + u_l(t_k)}{2} + c_1 + \varepsilon_1(t_k) \\ x_2(t_k) = \sum_{l=1}^d a_{2,l} \frac{y_l(t_{k-1}) + y_l(t_k)}{2} + \sum_{l=1}^p b_{2,l} \frac{u_l(t_{k-1}) + u_l(t_k)}{2} + c_2 + \varepsilon_2(t_k) \\ \vdots \\ x_d(t_k) = \sum_{l=1}^d a_{d,l} \frac{y_l(t_{k-1}) + y_l(t_k)}{2} + \sum_{l=1}^p b_{d,l} \frac{u_l(t_{k-1}) + u_l(t_k)}{2} + c_d + \varepsilon_d(t_k) \end{cases}$$

which can not only yield the existing multi-output models, such as the ones coupling self-memory theory [\[39\]](#) and multiple regression [\[53\]](#), but suggest other novel ones.

[Remarks 1–3](#) show that the unified [Eq. \(6\)](#) has the capacity to represent the continuous-time single-variable, multi-variable and multi-output grey models; and furthermore, it can also introduce the possibility of developing some other novel models.

### 3. Integral matching-based ordinary differential equation models

In this section we deduce the reduced-order representation of the aforementioned unified form; then the integral matching, which is composed of integral operator and least squares [\[54–56\]](#), is introduced to explain the mechanism of the cumulative sum operator as well as estimating the structural parameters and initial values simultaneously.

#### 3.1. Reduced-order representation of grey forecasting models

**Lemma 1.** Let the translation transformation of the Cusum series be  $\mathbf{y}_{tr}(t_k) = \mathbf{y}(t_k) + \boldsymbol{\xi}$ ,  $k = 1, 2, \dots, n$ , where  $\boldsymbol{\xi} \in \mathbb{R}^d$  is a real vector. Then, (i) the estimated parameters satisfy  $\hat{\mathbf{A}}_{tr} = \hat{\mathbf{A}}$ ,  $\hat{\mathbf{B}}_{tr} = \hat{\mathbf{B}}$ ,  $\hat{\mathbf{c}}_{tr} = \hat{\mathbf{c}} - \hat{\mathbf{A}}\boldsymbol{\xi}$ ; (ii) the time response functions satisfy  $\hat{\mathbf{y}}_{tr}(t) = \hat{\mathbf{y}}(t) + \boldsymbol{\xi}$ ; (iii) the fitting and forecasting values of the original time series satisfy  $\hat{\mathbf{x}}_{tr}(t_k) = \hat{\mathbf{x}}(t_k)$ ,  $k = 2, 3, \dots, n$ .

**Proof.** By substituting  $\mathbf{y}_{tr}(t_k) = \mathbf{y}(t_k) + \boldsymbol{\xi}$  into the matrices in [Eq. \(9\)](#), the two matrices therein are rewritten as

$$\mathbf{X}_{tr} = \mathbf{X} \text{ with } \mathbf{x}_{tr}(t_k) = \mathbf{x}(t_k) = \frac{1}{h_k} [\mathbf{y}_{tr}(t_k) - \mathbf{y}_{tr}(t_{k-1})], \quad k = 2, 3, \dots, n$$

and

$$\Theta(\mathbf{y}_{tr}, \mathbf{u}) = \begin{bmatrix} \frac{\mathbf{y}_{tr}^T(t_1) + \mathbf{y}_{tr}^T(t_2)}{2} & \frac{\mathbf{u}^T(t_1) + \mathbf{u}^T(t_2)}{2} & 1 \\ \frac{\mathbf{y}_{tr}^T(t_2) + \mathbf{y}_{tr}^T(t_3)}{2} & \frac{\mathbf{u}^T(t_2) + \mathbf{u}^T(t_3)}{2} & 1 \\ \vdots & \vdots & \vdots \\ \frac{\mathbf{y}_{tr}^T(t_{n-1}) + \mathbf{y}_{tr}^T(t_n)}{2} & \frac{\mathbf{u}^T(t_{n-1}) + \mathbf{u}^T(t_n)}{2} & 1 \end{bmatrix} = \Theta(\mathbf{y}, \mathbf{u}) \begin{bmatrix} \mathbf{I}_d & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p & \mathbf{0} \\ \boldsymbol{\xi}^T & \mathbf{0} & 1 \end{bmatrix}.$$



From Eq. (11), one has the estimated parameters expressed as

$$\begin{bmatrix} \hat{\mathbf{A}}_{\text{tr}}^{\top} \\ \hat{\mathbf{B}}_{\text{tr}}^{\top} \\ \hat{\mathbf{c}}_{\text{tr}}^{\top} \end{bmatrix} = (\Theta^{\top}(\mathbf{y}_{\text{tr}}, \mathbf{u})\Theta(\mathbf{y}_{\text{tr}}, \mathbf{u}))^{-1}\Theta^{\top}(\mathbf{y}_{\text{tr}}, \mathbf{u})\mathbf{X}_{\text{tr}} = \begin{bmatrix} \mathbf{I}_d & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p & \mathbf{0} \\ \boldsymbol{\xi}^{\top} & \mathbf{0} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{A}}^{\top} \\ \hat{\mathbf{B}}^{\top} \\ \hat{\mathbf{c}}^{\top} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{A}}^{\top} \\ \hat{\mathbf{B}}^{\top} \\ \hat{\mathbf{c}}^{\top} - \boldsymbol{\xi}^{\top}\hat{\mathbf{A}}^{\top} \end{bmatrix}.$$

Therefore, the corresponding time response function obtained from Eq. (12) is

$$\hat{\mathbf{y}}_{\text{tr}}(t) = \exp(\hat{\mathbf{A}}(t - t_1))\{\boldsymbol{\eta}_{\text{tr}} + \int_{t_1}^t \exp(\hat{\mathbf{A}}(t_1 - s))[\hat{\mathbf{B}}\mathbf{u}(s) + \hat{\mathbf{c}}]ds - \int_{t_1}^t \exp(\hat{\mathbf{A}}(t_1 - s))\hat{\mathbf{A}}\boldsymbol{\xi}ds\}$$

where, because both  $\boldsymbol{\eta}_{\text{tr}}$  and  $\boldsymbol{\eta}$  are determined by using the same strategy as in Eqs. (13), (14) or (15), the initial values satisfy

$$\boldsymbol{\eta}_{\text{tr}} = \boldsymbol{\eta} + \boldsymbol{\xi}$$

and the integral term at the right hand side is manipulated as

$$-\int_{t_1}^t \exp(\hat{\mathbf{A}}(t_1 - s))\hat{\mathbf{A}}\boldsymbol{\xi}ds = \exp(\hat{\mathbf{A}}(t_1 - t))\boldsymbol{\xi} - \boldsymbol{\xi}.$$

Finally, a direct substitution and simple manipulation give the time response function

$$\hat{\mathbf{y}}_{\text{tr}}(t) = \exp(\hat{\mathbf{A}}(t - t_1))\left\{\boldsymbol{\eta} + \int_{t_1}^t \exp(\hat{\mathbf{A}}(t_1 - s))[\hat{\mathbf{B}}\mathbf{u}(s) + \hat{\mathbf{c}}]ds\right\} + \boldsymbol{\xi} = \hat{\mathbf{y}}(t) + \boldsymbol{\xi}$$

and the restored fitting and forecasting values

$$\hat{\mathbf{x}}_{\text{tr}}(t_k) = \frac{\hat{\mathbf{y}}_{\text{tr}}(t_k) - \hat{\mathbf{y}}_{\text{tr}}(t_{k-1})}{t_k - t_{k-1}} = \frac{\hat{\mathbf{y}}(t_k) - \hat{\mathbf{y}}(t_{k-1})}{t_k - t_{k-1}} = \mathbf{x}(t_k), \quad k = 2, 3, \dots, n + r.$$

**Lemma 1** shows that the translation transformation of the Cusum series has no influence on the forecasting results. Note that the translation transformation of the Cusum series is equivalent to adding  $\boldsymbol{\xi}$  to the first element in the original time series  $(\mathbf{X}_{\text{tr}}(t) = \{\mathbf{x}(t_1) + \boldsymbol{\xi}, \mathbf{x}(t_2), \dots, \mathbf{x}(t_n)\})$ . Thus, without changing performance, the Cusum operator in Definition 1 can be defined also as  $\mathbf{y}(t_1) = \boldsymbol{\xi}$  and  $\mathbf{y}(t_k) = \boldsymbol{\xi} + \sum_{i=2}^k h_i \mathbf{x}(t_i)$  when  $k \geq 2$ .  $\square$

**Theorem 1.** Let

$$\mathbf{y}(t) = \boldsymbol{\xi} + \int_{t_1}^t \mathbf{x}(s)ds, \quad t \geq t_1 \quad (17)$$

where  $\boldsymbol{\xi} \in \mathbb{R}^d$  is a real vector. Then the ordinary differential Eq. (6) with initial value  $\mathbf{y}(t_1) = \boldsymbol{\xi}$  is equivalent to the reduced-order ordinary differential equation

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\frac{d}{dt}\mathbf{u}(t), \quad t \geq t_1 \quad (18)$$

with initial value  $\mathbf{x}(t_1) = \mathbf{A}\boldsymbol{\xi} + \mathbf{B}\mathbf{u}(t_1) + \mathbf{c}$ .

**Proof.** The proof is divided into two parts. Firstly, we prove that Eq. (6) can be reduced to Eq. (18). By substituting Eq. (17) into (6), one has

$$\mathbf{x}(t) = \mathbf{A}\left(\boldsymbol{\xi} + \int_{t_1}^t \mathbf{x}(s)ds\right) + \mathbf{B}\mathbf{u}(t) + \mathbf{c}, \quad t \geq t_1. \quad (19)$$

Differentiating both sides of Eq. (19) with respect to  $t$  gives the desired Eq. (18); then, combining Eq. (17) and the closed-form solution in Eq. (12), gives the initial value expressed as

$$\mathbf{x}(t_1) = \frac{d}{dt}\mathbf{y}(t)|_{t=t_1} = \mathbf{A}\boldsymbol{\xi} + \mathbf{B}\mathbf{u}(t_1) + \mathbf{c}.$$

Conversely, integrating both sides of Eq. (18) gives

$$\mathbf{x}(t) = \mathbf{A} \int_{t_1}^t \mathbf{x}(s)ds + \mathbf{B}\mathbf{u}(t) - \mathbf{B}\mathbf{u}(t_1) + \mathbf{x}(t_1), \quad t \geq t_1 \quad (20)$$

where, by differentiating both sides of Eq. (17) with respect to  $t$ , the left side is

$$\mathbf{x}(t) = \frac{d}{dt}\left(\boldsymbol{\xi} + \int_{t_1}^t \mathbf{x}(s)ds\right) = \frac{d}{dt}\mathbf{y}(t), \quad t \geq t_1$$

and by substituting  $\mathbf{y}(t) - \boldsymbol{\xi}$  in Eq. (17) for  $\int_{t_1}^t \mathbf{x}(s)ds$  in Eq. (20), the right side is

$$\mathbf{A} \int_{t_1}^t \mathbf{x}(s)ds + \mathbf{B}\mathbf{u}(t) - \mathbf{B}\mathbf{u}(t_1) + \mathbf{x}(t_1) = \mathbf{A}\mathbf{y}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{x}(t_1) - \mathbf{A}\boldsymbol{\xi} - \mathbf{B}\mathbf{u}(t_1).$$

Substituting the initial value  $\mathbf{x}(t_1) = \mathbf{c} + \mathbf{B}\mathbf{u}(t_1) + \mathbf{A}\boldsymbol{\xi}$  into Eq. (20) then yields the desired differential Eq. (6).  $\square$

**Theorem 1** shows that by introducing the integral operator (17), Eq. (6) can always be reduced to the simple form (18) and, in the other direction, Eq. (18) can express (6) by selecting an appropriate initial value  $\mathbf{x}(t_1) = \mathbf{c} + \mathbf{B}\mathbf{u}(t_1) + \mathbf{A}\boldsymbol{\xi}$ . In this sense, Eq. (6) has  $d$  redundant degrees of freedom (the parameter vector  $\mathbf{c}$ ) and it can be eliminated by using the reduced-order Eq. (18) to model the original time series directly. Here, we use an example to show the order reduction process. Supposing the ordinary differential Eq. (6) is  $\frac{d}{dt}y(t) = ay(t) + b_2t^2 + b_1t + c$ ,  $y(0) = \xi$ ,  $t \geq 0$ , whose closed-form solution is

$$y(t) = \left( \xi + \frac{c}{a} + \frac{b_1}{a^2} + \frac{2b_2}{a^3} \right) \exp(at) - \frac{b_2}{a}t^2 - \left( \frac{b_1}{a} + \frac{2b_2}{a^2} \right)t - \left( \frac{c}{a} + \frac{b_1}{a^2} + \frac{2b_2}{a^3} \right), \quad t \geq 0.$$

If  $y(t)$  satisfies Eq. (17), i.e.,  $y(t) = \xi + \int_0^t x(s)ds$ , then it is easy to derive that the equivalent reduced-order Eq. (18) is  $\frac{d}{dt}x(t) = ax(t) + 2b_2t + b_1$ ,  $x(0) = a\xi + c$ ,  $t \geq 0$ , whose closed-form solution is

$$x(t) = \left( a\xi + c + \frac{b_1}{a} + \frac{2b_2}{a^2} \right) \exp(at) - \frac{2b_2}{a}t - \left( \frac{b_1}{a} + \frac{2b_2}{a^2} \right), \quad t \geq 0.$$

Since Lemma 1 shows that  $\boldsymbol{\xi}$  has no influence on modelling performance then, without loss of generality, let  $\boldsymbol{\xi} = \mathbf{x}(t_1)$  in order to be consistent with the intuitive understanding of the Cusum operator. In this way, the initial values satisfy  $\mathbf{y}(t_1) = \boldsymbol{\xi} = \mathbf{x}(t_1)$  and the redundant parameter vector satisfy  $\mathbf{c} = (\mathbf{I} - \mathbf{A})\mathbf{y}(t_1) - \mathbf{B}\mathbf{u}(t_1)$ . This provides an alternative strategy for initial value selection, i.e.,  $\hat{\mathbf{y}}(t_1) = (\mathbf{I} - \hat{\mathbf{A}})^{-1}(\hat{\mathbf{c}} + \hat{\mathbf{B}}\mathbf{u}(t_1))$ .

### 3.2. Discretization of integral operator for explaining the cumulative sum operator

The integral operator (17) plays an important role in the simplification from Eqs. (6) to (18). Since the observations of the state vector  $\mathbf{x}(t)$  at time points  $\{t_k\}_{k=1}^n$  are known in advance, the finite integral in Eq. (17) can be approximated using the piecewise interpolation methods, such as the piecewise constant integration formula

$$\mathbf{y}_{\text{cnt}}(t_k) = \mathbf{x}(t_1) + \int_{t_1}^t \mathbf{x}(s)ds = \begin{cases} \mathbf{x}(t_1), & k = 1 \\ \mathbf{x}(t_1) + \sum_{i=2}^k h_i \mathbf{x}(t_i), & k = 2, 3, \dots, n \end{cases} \quad (21)$$

and the piecewise linear integration formula

$$\mathbf{y}_{\text{int}}(t_k) = \mathbf{x}(t_1) + \int_{t_1}^t \mathbf{x}(s)ds = \begin{cases} \mathbf{x}(t_1), & k = 1 \\ \mathbf{x}(t_1) + \frac{1}{2} \sum_{i=2}^k h_i \mathbf{x}(t_{i-1}) + \frac{1}{2} \sum_{i=2}^k h_i \mathbf{x}(t_i), & k = 2, 3, \dots, n \end{cases} \quad (22)$$

**Remark 4.** The piecewise constant integral formula in Eq. (21) is equivalent to the Cusum operator in Definition 1, indicating that the Cusum operator is a numerical discretization of the integral operator (17) and, in turn, the latter is the continuous generalization of the former.

Note that Eq. (22) is a second-order formula and thus has higher accuracy than the first-order (21). Furthermore, other higher-order numerical integration approaches can also be used to further improve the accuracy when calculating the finite integration in Eq. (17), such as the Simpson rule [57].

### 3.3. Integral matching for simultaneous structural parameter and initial value estimation

Integral matching-based model uses the reduced-order Eq. (18) to fit the original time series directly, and the key is to estimate the structural parameter and initial value from sampled time-series data.

**Theorem 1** shows that integrating both sides of Eq. (18) gives the definite integral Eq. (20). By substituting  $\mathbf{x}(t_1) = \boldsymbol{\eta}$  into Eq. (20), one has

$$\mathbf{x}(t) = \mathbf{A} \int_{t_1}^t \mathbf{x}(s)ds + \mathbf{B}(\mathbf{u}(t) - \mathbf{u}(t_1)) + \boldsymbol{\eta}, \quad t \geq t_1;$$

and then, using the numerical integration formula (22) to approximate the integrals, we obtain

$$\mathbf{x}(t_k) = \mathbf{A}[\mathbf{y}_{\text{int}}(t_k) - \mathbf{x}(t_1)] + \mathbf{B}[\mathbf{u}(t_k) - \mathbf{u}(t_1)] + \boldsymbol{\eta} + \boldsymbol{\epsilon}(t_k) \quad (23)$$

where  $\boldsymbol{\epsilon}(t_k)$  is the sum of numerical error and noise error at each time instant.

Similar to the treatment process in Section 2, by substituting  $k = 2, 3, \dots, n$  and the observed data into Eq. (23), and then arranging the  $n - 1$  algebraic equations into a matrix form, one has

$$\mathbf{X} = \boldsymbol{\Omega}(\mathbf{x}, \mathbf{u})\boldsymbol{\Pi} + \boldsymbol{\Upsilon} \quad (24)$$

where

$$\Pi = \begin{bmatrix} \mathbf{A}^\top \\ \mathbf{B}^\top \\ \boldsymbol{\eta}^\top \end{bmatrix}, \quad \Omega(\mathbf{x}, \mathbf{u}) = \begin{bmatrix} \mathbf{y}_{\text{Int}}^\top(t_2) - \mathbf{x}^\top(t_1) & \mathbf{u}^\top(t_2) - \mathbf{u}^\top(t_1) & 1 \\ \mathbf{y}_{\text{Int}}^\top(t_3) - \mathbf{x}^\top(t_1) & \mathbf{u}^\top(t_3) - \mathbf{u}^\top(t_1) & 1 \\ \vdots & \vdots & \vdots \\ \mathbf{y}_{\text{Int}}^\top(t_n) - \mathbf{x}^\top(t_1) & \mathbf{u}^\top(t_n) - \mathbf{u}^\top(t_1) & 1 \end{bmatrix}, \quad \Upsilon = \begin{bmatrix} \boldsymbol{\epsilon}^\top(t_2) \\ \boldsymbol{\epsilon}^\top(t_3) \\ \vdots \\ \boldsymbol{\epsilon}^\top(t_n) \end{bmatrix}.$$

And by minimizing the objective function  $\mathcal{L}(\Pi) = \|\mathbf{X} - \Omega(\mathbf{x}, \mathbf{u})\Pi\|_F^2$ , one has the least-square parameter matrix

$$\hat{\Pi} = (\Omega^\top(\mathbf{x}, \mathbf{u})\Omega(\mathbf{x}, \mathbf{u}))^{-1}\Omega^\top(\mathbf{x}, \mathbf{u})\mathbf{X}. \quad (25)$$

Substituting the estimates of the structural parameters ( $\hat{\mathbf{A}}$  and  $\hat{\mathbf{B}}$ ) and initial value ( $\hat{\boldsymbol{\eta}}$ ) into the general solution to the reduced-order Eq. (18), gives the time response function

$$\hat{\mathbf{x}}(t) = \exp(\hat{\mathbf{A}}(t - t_1))\{\hat{\boldsymbol{\eta}} + \exp(\hat{\mathbf{A}}(t_1 - t))\hat{\mathbf{B}}\mathbf{u}(t) - \hat{\mathbf{B}}\mathbf{u}(t_1) + \int_{t_1}^t \exp(\hat{\mathbf{A}}(t_1 - s))\hat{\mathbf{A}}\hat{\mathbf{B}}\mathbf{u}(s)ds\} \quad (26)$$

and then the fitting and forecasting values of the original time series can be obtained by substituting the time points  $\{t_k\}_{k=1}^{n+r}$  into Eq. (26).

#### 4. The relationship between grey models and integral matching-based models

In this section we analyse the relationship between grey forecasting models and integral matching-based differential equation models from both the parameter estimation and the modelling procedures perspectives. For convenience, the original time series is assumed to be equally spaced, i.e.,  $h_k = h$ ,  $k = 2, 3, \dots, n$ . The estimated parameters of grey models are denoted as  $\hat{\mathbf{A}}_g$ ,  $\hat{\mathbf{B}}_g$  and  $\hat{\mathbf{c}}_g$ , and those of integral matching-based ones are denoted as  $\hat{\mathbf{A}}_m$ ,  $\hat{\mathbf{B}}_m$  and  $\hat{\boldsymbol{\eta}}_m$ .

##### 4.1. Quantitative relationship between the estimated parameters

**Proposition 1.** Let the independent vector be  $\mathbf{u}(t) = 0$ . The estimated parameters satisfy  $\hat{\mathbf{A}}_g = \hat{\mathbf{A}}_m$  and  $\hat{\boldsymbol{\eta}}_m = \hat{\mathbf{c}}_g + \hat{\mathbf{A}}_g\mathbf{x}(t_1) - \frac{h}{2}\hat{\mathbf{A}}_g\mathbf{x}(t_1)$ .

**Proof.** From Eqs. (11) and (25), it is easy to show that the estimated parameters are

$$\begin{bmatrix} \hat{\mathbf{A}}_g^\top \\ \hat{\mathbf{c}}_g^\top \end{bmatrix} = (\mathbf{Y}_g^\top \mathbf{Y}_g)^{-1} \mathbf{Y}_g^\top \mathbf{X} \quad \text{and} \quad \begin{bmatrix} \hat{\mathbf{A}}_m^\top \\ \hat{\boldsymbol{\eta}}_m^\top \end{bmatrix} = (\mathbf{Y}_m^\top \mathbf{Y}_m)^{-1} \mathbf{Y}_m^\top \mathbf{X}$$

where

$$\mathbf{Y}_g = \begin{bmatrix} \frac{\mathbf{y}^\top(t_1) + \mathbf{y}^\top(t_2)}{2} & 1 \\ \frac{\mathbf{y}^\top(t_2) + \mathbf{y}^\top(t_3)}{2} & 1 \\ \vdots & \vdots \\ \frac{\mathbf{y}^\top(t_{n-1}) + \mathbf{y}^\top(t_n)}{2} & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{Y}_m = \begin{bmatrix} \mathbf{y}_{\text{Int}}^\top(t_2) - \mathbf{x}^\top(t_1) & 1 \\ \mathbf{y}_{\text{Int}}^\top(t_3) - \mathbf{x}^\top(t_1) & 1 \\ \vdots & \vdots \\ \mathbf{y}_{\text{Int}}^\top(t_n) - \mathbf{x}^\top(t_1) & 1 \end{bmatrix}.$$

Because the original time series is equally spaced, Eq. (22) can be rewritten as

$$\mathbf{y}_{\text{Int}}(t_k) = \frac{1}{2}\mathbf{y}(t_{k-1}) + \frac{1}{2}\mathbf{y}(t_k) + \frac{h}{2}\mathbf{x}(t_1), \quad k = 2, 3, \dots, n$$

and thus we have  $\mathbf{Y}_m = \mathbf{Y}_g \mathbf{Q}$ , where  $\mathbf{I}_d$  is the  $d$ th-order unit matrix,

$$\mathbf{Q} = \begin{bmatrix} \mathbf{I}_d & 0 \\ \frac{h-2}{2}\mathbf{x}^\top(t_1) & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{Q}^{-1} = \begin{bmatrix} \mathbf{I}_d & 0 \\ -\frac{h-2}{2}\mathbf{x}^\top(t_1) & 1 \end{bmatrix}$$

Finally, a simple substitution gives the desired equation

$$\begin{bmatrix} \hat{\mathbf{A}}_m^\top \\ \hat{\boldsymbol{\eta}}_m^\top \end{bmatrix} = (\mathbf{Y}_g \mathbf{Q})^\top (\mathbf{Y}_g \mathbf{Q})^{-1} (\mathbf{Y}_g \mathbf{Q})^\top \mathbf{X} = \mathbf{Q}^{-1} \begin{bmatrix} \hat{\mathbf{A}}_g^\top \\ \hat{\mathbf{c}}_g^\top \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{A}}_g^\top \\ \hat{\mathbf{c}}_g^\top + \mathbf{x}^\top(t_1)\hat{\mathbf{A}}_g^\top - \frac{h}{2}\mathbf{x}^\top(t_1)\hat{\mathbf{A}}_g^\top \end{bmatrix}.$$

Proposition 1 shows that the estimated matrices satisfy  $\hat{\mathbf{A}}_g = \hat{\mathbf{A}}_m$  and  $\hat{\mathbf{c}}_g = \hat{\boldsymbol{\eta}}_m - \hat{\mathbf{A}}_g\mathbf{x}(t_1) + \frac{h}{2}\hat{\mathbf{A}}_g\mathbf{x}(t_1)$ , showing that grey models can be viewed as equivalent forms of integral matching-based models by applying a translation transformation (translation coefficient is  $\xi = \frac{h-2}{2}\mathbf{x}(t_1)$  in Lemma 1) to the Cusum series, and considering the impact of translation transformation on the estimates is  $\frac{2-h}{2}\hat{\mathbf{A}}_g\mathbf{x}(t_1)$  in  $\hat{\mathbf{c}}_g$ .  $\square$

If the independent vector  $\mathbf{u}(t) \neq 0$ , the quantitative relationship is not that obvious. By partitioning the estimated parameter matrices in Eqs. (11) and (25) as  $\hat{\Sigma}_g = [\hat{\mathbf{A}}_g \ \hat{\mathbf{c}}_g \mid \hat{\mathbf{B}}_g]$  and  $\hat{\Pi}_m = [\hat{\mathbf{A}}_m \ \hat{\eta}_m \mid \hat{\mathbf{B}}_m]$ , then we have

$$\hat{\Sigma}_g = \left( \begin{bmatrix} \mathbf{Y}_g^\top \\ \mathbf{U}_g^\top \end{bmatrix} [\mathbf{Y}_g \ \mathbf{U}_g] \right)^{-1} \begin{bmatrix} \mathbf{Y}_g^\top \\ \mathbf{U}_g^\top \end{bmatrix} \mathbf{X} = \begin{bmatrix} \mathbf{Y}_g^\top \mathbf{Y}_g & \mathbf{Y}_g^\top \mathbf{U}_g \\ \mathbf{U}_g^\top \mathbf{Y}_g & \mathbf{U}_g^\top \mathbf{U}_g \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Y}_g^\top \mathbf{X} \\ \mathbf{U}_g^\top \mathbf{X} \end{bmatrix} \quad (27)$$

and

$$\hat{\Pi}_m = \left( \begin{bmatrix} \mathbf{Y}_m^\top \\ \mathbf{U}_m^\top \end{bmatrix} [\mathbf{Y}_m \ \mathbf{U}_m] \right)^{-1} \begin{bmatrix} \mathbf{Y}_m^\top \\ \mathbf{U}_m^\top \end{bmatrix} \mathbf{X} = \begin{bmatrix} \mathbf{Y}_m^\top \mathbf{Y}_m & \mathbf{Y}_m^\top \mathbf{U}_m \\ \mathbf{U}_m^\top \mathbf{Y}_m & \mathbf{U}_m^\top \mathbf{U}_m \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Y}_m^\top \mathbf{X} \\ \mathbf{U}_m^\top \mathbf{X} \end{bmatrix} \quad (28)$$

where

$$\mathbf{U}_g = \frac{1}{2} \begin{bmatrix} \mathbf{u}^\top(t_2) + \mathbf{u}^\top(t_1) \\ \mathbf{u}^\top(t_3) + \mathbf{u}^\top(t_2) \\ \vdots \\ \mathbf{u}^\top(t_n) + \mathbf{u}^\top(t_{n-1}) \end{bmatrix} = \begin{bmatrix} \mathbf{u}^\top(t_2) \\ \mathbf{u}^\top(t_3) \\ \vdots \\ \mathbf{u}^\top(t_n) \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \mathbf{u}^\top(t_1) - \mathbf{u}^\top(t_2) \\ \mathbf{u}^\top(t_2) - \mathbf{u}^\top(t_3) \\ \vdots \\ \mathbf{u}^\top(t_{n-1}) - \mathbf{u}^\top(t_n) \end{bmatrix}$$

and

$$\mathbf{U}_m = \begin{bmatrix} \mathbf{u}^\top(t_2) - \mathbf{u}^\top(t_1) \\ \mathbf{u}^\top(t_3) - \mathbf{u}^\top(t_2) \\ \vdots \\ \mathbf{u}^\top(t_n) - \mathbf{u}^\top(t_{n-1}) \end{bmatrix} = \begin{bmatrix} \mathbf{u}^\top(t_2) \\ \mathbf{u}^\top(t_3) \\ \vdots \\ \mathbf{u}^\top(t_n) \end{bmatrix} - \begin{bmatrix} \mathbf{u}^\top(t_1) \\ \mathbf{u}^\top(t_2) \\ \vdots \\ \mathbf{u}^\top(t_{n-1}) \end{bmatrix} = \mathbf{U}_g + \frac{1}{2} \begin{bmatrix} \mathbf{u}^\top(t_2) - \mathbf{u}^\top(t_1) \\ \mathbf{u}^\top(t_3) - \mathbf{u}^\top(t_2) \\ \vdots \\ \mathbf{u}^\top(t_n) - \mathbf{u}^\top(t_{n-1}) \end{bmatrix} - \begin{bmatrix} \mathbf{u}^\top(t_1) \\ \mathbf{u}^\top(t_2) \\ \vdots \\ \mathbf{u}^\top(t_{n-1}) \end{bmatrix}.$$

By denoting  $\mathbf{U}_m = \mathbf{U}_g + \mathbf{U}$  and combining  $\mathbf{Y}_m = \mathbf{Y}_g \mathbf{Q}$  in Proposition 1, then the matrices in Eq. (28) can be rewritten as

$$\begin{bmatrix} \mathbf{Y}_m^\top \mathbf{X} \\ \mathbf{U}_m^\top \mathbf{X} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix} \left( \begin{bmatrix} \mathbf{Y}_g^\top \mathbf{X} \\ \mathbf{U}_g^\top \mathbf{X} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{U}^\top \mathbf{X} \end{bmatrix} \right) \quad (29)$$

and

$$\begin{bmatrix} \mathbf{Y}_m^\top \mathbf{Y}_m & \mathbf{Y}_m^\top \mathbf{U}_m \\ \mathbf{U}_m^\top \mathbf{Y}_m & \mathbf{U}_m^\top \mathbf{U}_m \end{bmatrix} = \begin{bmatrix} \mathbf{Q}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix} \left( \begin{bmatrix} \mathbf{Y}_g^\top \mathbf{Y}_g & \mathbf{Y}_g^\top \mathbf{U}_g \\ \mathbf{U}_g^\top \mathbf{Y}_g & \mathbf{U}_g^\top \mathbf{U}_g \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{Y}_g^\top \mathbf{U} \\ \mathbf{U}^\top \mathbf{Y}_g & \mathbf{W} \end{bmatrix} \right) \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix} \quad (30)$$

where  $\mathbf{W} = \mathbf{U}_g^\top \mathbf{U} + \mathbf{U}^\top \mathbf{U}_g + \mathbf{U}^\top \mathbf{U}$ .

By substituting Eqs. (29) and (30) into Eq. (28), the estimates are obtained as

$$\hat{\Pi}_m = \begin{bmatrix} \mathbf{Q}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix} \left( \begin{bmatrix} \mathbf{Y}_g^\top \mathbf{Y}_g & \mathbf{Y}_g^\top \mathbf{U}_g \\ \mathbf{U}_g^\top \mathbf{Y}_g & \mathbf{U}_g^\top \mathbf{U}_g \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{Y}_g^\top \mathbf{U} \\ \mathbf{U}^\top \mathbf{Y}_g & \mathbf{W} \end{bmatrix} \right)^{-1} \left( \begin{bmatrix} \mathbf{Y}_g^\top \mathbf{X} \\ \mathbf{U}_g^\top \mathbf{X} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{U}^\top \mathbf{X} \end{bmatrix} \right) \quad (31)$$

where, by using the Sherman-Morrison-Woodbury formula [58], the inverse matrix can be rewritten as

$$\begin{bmatrix} \mathbf{Y}_g^\top \mathbf{Y}_g & \mathbf{Y}_g^\top \mathbf{U}_g \\ \mathbf{U}_g^\top \mathbf{Y}_g & \mathbf{U}_g^\top \mathbf{U}_g \end{bmatrix}^{-1} - \begin{bmatrix} \mathbf{Y}_g^\top \mathbf{Y}_g & \mathbf{Y}_g^\top \mathbf{U}_g \\ \mathbf{U}_g^\top \mathbf{Y}_g & \mathbf{U}_g^\top \mathbf{U}_g \end{bmatrix}^{-1} \left( \begin{bmatrix} \mathbf{Y}_g^\top \mathbf{Y}_g & \mathbf{Y}_g^\top \mathbf{U}_g \\ \mathbf{U}_g^\top \mathbf{Y}_g & \mathbf{U}_g^\top \mathbf{U}_g \end{bmatrix}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{Y}_g^\top \mathbf{U} \\ \mathbf{U}^\top \mathbf{Y}_g & \mathbf{W} \end{bmatrix}^{-1} \right) \begin{bmatrix} \mathbf{Y}_g^\top \mathbf{Y}_g & \mathbf{Y}_g^\top \mathbf{U}_g \\ \mathbf{U}_g^\top \mathbf{Y}_g & \mathbf{U}_g^\top \mathbf{U}_g \end{bmatrix}^{-1}.$$

Therefore, it follows that

$$\begin{aligned} \hat{\Pi}_m &= \begin{bmatrix} \mathbf{Q}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix} \hat{\Sigma}_g \\ &+ \begin{bmatrix} \mathbf{Q}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix} \begin{bmatrix} \mathbf{Y}_g^\top \mathbf{Y}_g & \mathbf{Y}_g^\top \mathbf{U}_g \\ \mathbf{U}_g^\top \mathbf{Y}_g & \mathbf{U}_g^\top \mathbf{U}_g \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{U}^\top \mathbf{X} \end{bmatrix} \\ &- \begin{bmatrix} \mathbf{Q}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix} \begin{bmatrix} \mathbf{Y}_g^\top \mathbf{Y}_g & \mathbf{Y}_g^\top \mathbf{U}_g \\ \mathbf{U}_g^\top \mathbf{Y}_g & \mathbf{U}_g^\top \mathbf{U}_g \end{bmatrix}^{-1} \left( \begin{bmatrix} \mathbf{Y}_g^\top \mathbf{Y}_g & \mathbf{Y}_g^\top \mathbf{U}_g \\ \mathbf{U}_g^\top \mathbf{Y}_g & \mathbf{U}_g^\top \mathbf{U}_g \end{bmatrix}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{Y}_g^\top \mathbf{U} \\ \mathbf{U}^\top \mathbf{Y}_g & \mathbf{W} \end{bmatrix}^{-1} \right) \hat{\Sigma}_g \\ &- \begin{bmatrix} \mathbf{Q}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix} \begin{bmatrix} \mathbf{Y}_g^\top \mathbf{Y}_g & \mathbf{Y}_g^\top \mathbf{U}_g \\ \mathbf{U}_g^\top \mathbf{Y}_g & \mathbf{U}_g^\top \mathbf{U}_g \end{bmatrix}^{-1} \left( \begin{bmatrix} \mathbf{Y}_g^\top \mathbf{Y}_g & \mathbf{Y}_g^\top \mathbf{U}_g \\ \mathbf{U}_g^\top \mathbf{Y}_g & \mathbf{U}_g^\top \mathbf{U}_g \end{bmatrix}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{Y}_g^\top \mathbf{U} \\ \mathbf{U}^\top \mathbf{Y}_g & \mathbf{W} \end{bmatrix}^{-1} \right) \begin{bmatrix} \mathbf{Y}_g^\top \mathbf{Y}_g & \mathbf{Y}_g^\top \mathbf{U}_g \\ \mathbf{U}_g^\top \mathbf{Y}_g & \mathbf{U}_g^\top \mathbf{U}_g \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{U}^\top \mathbf{X} \end{bmatrix} \end{aligned}$$

where the first term on the right hand side shows that

$$\hat{\mathbf{A}}_g = \hat{\mathbf{A}}_m, \ \hat{\mathbf{B}}_g = \hat{\mathbf{B}}_m, \ \hat{\eta}_m = \hat{\mathbf{c}}_g - \frac{h-2}{2} \hat{\mathbf{A}}_g \mathbf{x}(t_1)$$

and the other three terms can be viewed as the calibration of the estimated parameter matrices by using the information in the matrix  $\mathbf{U}$ . In particular, the solution in Eq. (31) yields that in Proposition 1 when the matrix is  $\mathbf{U} = \mathbf{0}$ .

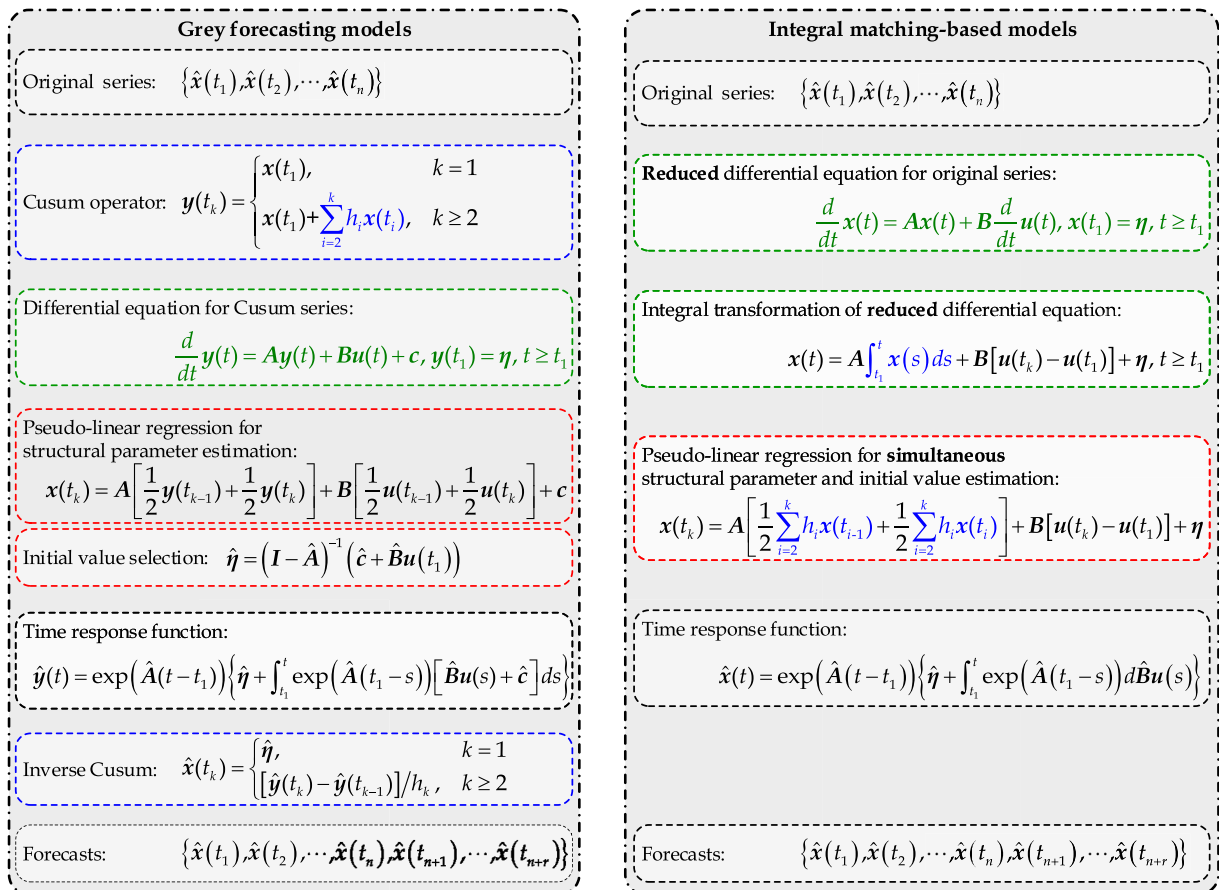


Fig. 2. Reconstructed modelling procedures of grey models by integral matching-based models.

#### 4.2. Modelling procedures reconstruction of grey forecasting models

The main objective of both grey models and integral matching-based models is the explanation of time-series data using ordinary differential equations. The modelling procedures of both models are summarized in Figure 2.

Fig. 2 shows that the modelling procedures of integral matching-based models can be viewed as the simplification and reconstruction of grey models. In particular:

- (i) Grey models use ordinary differential equations to fit the Cusum series, while integral matching-based models employ reduced-order differential equations to explain and approximate the original series.
- (ii) Both of them use the trapezoid rule to discretize the ordinary differential equations into the discrete-time pseudo-regression expressions when estimating the parameters.
- (iii) The integral matching-based models obtain the estimates of structural parameters and initial values simultaneously, whereas the conventional grey models need to select a strategy to determine the initial value, such as the least-square strategy in Eq. (15), which may well introduce additional errors [59].
- (iv) Both of them use the time response function to compute the fitting and forecasting values. The grey models need to use the inverse Cusum operator to restore the forecasting results of the Cusum series, which makes this kind of model more tedious than the integral matching-based ones.

The above modelling procedures show that both grey and integral matching-based models are constructed under the assumption that each independent variable  $u_i(t)$  is known in advance, such as the deterministic function of time in Remark 1. In some cases, there exists time-lag effect between the state variable and independent variable, such as the cases in Remark 2 having  $u_i(t) = y_i(t - \tau_i h)$  where  $\tau_i \geq 1$  is the time delay order and  $h$  is the time interval. It is worth noting, however, that the possibility of complex 'forcing variables' occurs often in science and economics. In forecasting terms, such complex variables need to be accounted for as 'leading indicators' over the forecasting horizon. In this situation, these input variables will themselves have to be modelled and forecasted over the forecasting horizon. For example, univariate or multi-

variate time series methods can be applied, such as the dynamic harmonic regression [60,61]; see [62] for a comprehensive example.

## 5. Simulations

In this section large-scale Monte Carlo simulations are conducted to compare grey models and integral matching-based models from both parameter estimation and forecasting accuracy perspectives.

The time-series data are generated from the observation equation

$$\tilde{\mathbf{x}}(t_k) = \begin{cases} \mathbf{x}(t_k) + \mathbf{e}(k), & k = 1, 2, \dots, n \quad \leftarrow \text{in-sample series} \\ \mathbf{x}(t_k), & k = n + 1, \dots, n + r \quad \leftarrow \text{out-of-sample series} \end{cases} \quad (32)$$

where  $\mathbf{x}(t)$  is the solution to the state Eq. (18);  $\mathbf{e}(k) \sim \mathcal{N}(\mathbf{0}, \Sigma)$  is a  $d$ -dimensional normal distribution with mean  $\mathbf{0}$  and covariance matrix  $\Sigma = \sigma^2 \mathbf{I}_d$ . The fitting and forecasting errors are measured by the in-sample and out-of-sample mean absolute percentage error criteria, respectively expressed as

$$\text{MAPE}_{\text{in}}[x_\ell] = \frac{1}{n} \sum_{k=1}^n \left| \frac{\hat{x}_\ell(t_k) - \tilde{x}_\ell(t_k)}{\tilde{x}_\ell(t_k)} \right| \times 100\% \text{ and } \text{MAPE}_{\text{out}}[x_\ell] = \frac{1}{r} \sum_{k=n+1}^{n+r} \left| \frac{\hat{x}_\ell(t_k) - x_\ell(t_k)}{x_\ell(t_k)} \right| \times 100\%$$

where  $x_\ell(t)$ ,  $\ell = 1, 2, \dots, d$ , is the  $\ell$ -th component of the state variable  $\mathbf{x}(t)$ .

### 5.1. Data generating process

To analyse the effect of sample size and noise level on modelling results, the sample size  $n$  and standard deviation  $\sigma$  are set to different values. Here, we select a 2-dimensional autonomous system as a comprehensive example. Simulations of single-variable and other multi-variable models can be conducted by following similar steps. Let the state Eq. (18) be

$$\begin{cases} \frac{d}{dt} x_1(t) = -0.25x_1(t) + 0.70x_2(t) \\ \frac{d}{dt} x_2(t) = 0.75x_1(t) - 0.25x_2(t) \end{cases}, \quad \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} 1.20 \\ 0.35 \end{bmatrix}, \quad t \in [0, 5]$$

which indicates that the true values of structural parameter matrix and initial vector are

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} = \begin{bmatrix} -0.25 & 0.70 \\ 0.75 & -0.25 \end{bmatrix}, \quad \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 1.20 \\ 0.35 \end{bmatrix}.$$

Taking the time interval as  $h \in [0.25, 0.10, 0.05]$ , then the observations are sampled at every time interval of  $h$  in the range of  $t \in [0, 5]$ , thereby generating  $n = \left\lceil \frac{5}{h} \right\rceil + 1 \in \{21, 51, 101\}$  samples; using the signal-to-noise ratio  $\text{snr} \in \{2.5, 3.5, 5.0\}$  to measure the noise level, then the standard deviation is  $\sigma = \frac{1}{\sqrt{\text{snr}}} \sqrt{\text{Var}(x_\ell(t))}$ . By substituting  $n$  and  $\sigma$  into Eq. (32), a total of  $3 \times 3 = 9$  scenarios are obtained and in each sample size and signal-to-noise ratio combination scenario, 1000 realizations are replicated.

Before reporting the modelling results, it should be noted from Theorem 1 that the equivalent equation utilized in grey model (20) is

$$\begin{cases} \frac{d}{dt} y_1(t) = -0.25y_1(t) + 0.70y_2(t) + 1.2550 \\ \frac{d}{dt} y_2(t) = 0.75y_1(t) - 0.25y_2(t) - 0.4625 \end{cases}, \quad \begin{bmatrix} y_1(0) \\ y_2(0) \end{bmatrix} = \begin{bmatrix} 1.20 \\ 0.35 \end{bmatrix}, \quad t \in [0, 5].$$

### 5.2. Evaluation of parameter estimation performance

The estimations of structural parameters and initial vectors obtained from grey modelling and integral matching techniques are summarized in Table 2. Here, the structural parameter estimation matrices obtained from these two approaches are identical, validating Proposition 1.

On the basis of the results in Table 2, we may roughly conclude the following: the estimates of parameters, including structural parameters and initial values, tend to their corresponding true values, and meanwhile, the sample standard deviations of estimates get smaller with the increase of sample size, partially implying the asymptotic behaviour of these two approaches; the estimates of the initial vector obtained from integral matching are superior to those obtained from grey modelling in sample mean and sample standard deviation terms, indicating that integral matching has higher accuracy and robustness. To be specific, assuming that the signal-to-noise ratio remains fixed, then the 9 scenarios are reshaped to be 3 groups; and in each group, as the sample size increases, the sample standard deviations of the estimates reduce successively, validating the efficiency of the estimators; whereas the sample means of the estimates do not get closer to their true values monotonically with the increase of sample size, indicating the likely existence of asymptotic bias on the estimates.

**Table 2**

Sample means (sample standard deviations) of the estimated parameters obtained from grey modelling and integral matching approaches in 1000 runs with the sample size  $n = 21, 51, 101$  and signal-to-noise ratio  $snr = 2.5, 3.5, 5.0$ . The true parameters are  $a_{1,1} = -0.25$ ,  $a_{1,2} = 0.70$ ,  $a_{2,1} = 0.75$ ,  $a_{2,2} = -0.25$ ,  $\eta_1 = 1.20$  and  $\eta_2 = 0.35$ .

$n$	$snr$	Structural parameter matrix <sup>a</sup>				Initial vector <sup>b</sup>		Initial vector <sup>c</sup>	
		$\hat{a}_{1,1}$	$\hat{a}_{1,2}$	$\hat{a}_{2,1}$	$\hat{a}_{2,2}$	$\hat{\eta}_1$	$\hat{\eta}_2$	$\hat{\eta}_1$	$\hat{\eta}_2$
21	2.5	-0.249(0.606)	0.699(0.601)	0.745(0.655)	-0.245(0.651)	1.301(0.681)	0.509(0.711)	1.195(0.330)	0.346(0.368)
	3.5	-0.251(0.309)	0.701(0.306)	0.743(0.334)	-0.244(0.332)	1.278(0.343)	0.486(0.358)	1.201(0.168)	0.351(0.188)
	5.0	-0.250(0.151)	0.700(0.150)	0.744(0.164)	-0.245(0.162)	1.268(0.168)	0.476(0.175)	1.201(0.082)	0.351(0.092)
51	2.5	-0.255(0.342)	0.706(0.341)	0.742(0.371)	-0.241(0.370)	1.252(0.452)	0.426(0.465)	1.200(0.175)	0.350(0.197)
	3.5	-0.254(0.174)	0.704(0.174)	0.744(0.189)	-0.244(0.188)	1.238(0.229)	0.411(0.236)	1.201(0.089)	0.351(0.100)
	5.0	-0.252(0.085)	0.702(0.085)	0.747(0.092)	-0.247(0.092)	1.231(0.112)	0.404(0.115)	1.201(0.044)	0.351(0.049)
101	2.5	-0.236(0.224)	0.686(0.223)	0.764(0.242)	-0.264(0.241)	1.223(0.371)	0.384(0.377)	1.194(0.115)	0.343(0.129)
	3.5	-0.244(0.115)	0.694(0.114)	0.756(0.124)	-0.257(0.123)	1.218(0.189)	0.379(0.192)	1.197(0.059)	0.347(0.066)
	5.0	-0.247(0.056)	0.697(0.056)	0.753(0.061)	-0.253(0.060)	1.215(0.093)	0.376(0.094)	1.199(0.029)	0.349(0.032)

<sup>a</sup> The estimated structural parameter matrices obtained from grey modelling and integral matching are identical.

<sup>b</sup> The estimated initial vector of grey modelling is obtained from  $\hat{\eta} = (I_2 - \hat{A})^{-1}\hat{c}$ .

<sup>c</sup> The estimated initial vector corresponds to integral matching.

### 5.3. Evaluation of fitting and forecasting accuracy

Since grey models are widely used to forecast time series, it is also necessary to analyse the distributions of the fitting and forecasting errors. Fig. 3 depicts the boxplots of the fitting errors, two-step-ahead, five-step-ahead, and ten-step-ahead forecasting errors in 1000 simulations.

As a whole, for both the first component  $x_1$  in Fig. 3(a) and the second component  $x_2$  in Fig. 3(b), integral matching has much smaller fitting and multi-step-ahead forecasting errors than grey modelling in all the sample size and signal-to-noise ratio combination scenarios, indicating that the former outperforms the latter from both fitting and forecasting viewpoints. Specifically, by the same token as the grouping manner in Section 5.2, we may also conclude the following:

- For each group with an equal sample size, i.e., each row in Fig. 3(a) and (b), the fitting and multi-step-ahead forecasting errors decrease successively with the increase of signal-to-noise ratio, indicating the impact of measurement noise on the forecasting accuracy. Let's take the example of the integral matching method. Letting  $n = 21$ , then, as  $snr$  increases from 2.5 through 3.5 to 5.0, the medians of fitting errors decreases from 6.34% through 3.22% to 1.58% (for  $x_1$ ) and from 11.57% through 5.89% to 2.89% (for  $x_2$ ), and meanwhile, the medians of the ten-step-ahead forecasting errors decreases from 6.63% through 3.24% to 1.59% (for  $x_1$ ) and from 6.49% through 3.27% to 1.61% (for  $x_2$ ).
- For each group with an equal signal-to-noise ratio, i.e., each column in Fig. 3(a) and (b), as the sample size increases, the fitting and multi-step-ahead forecasting errors decrease monotonously, partially implying the asymptotic unbiasedness of forecasts. Likewise, we still use the integral matching method as an example. Letting  $snr = 2.5$ , then, as  $n$  increases from 21 through 51 to 101, the medians of fitting errors decreases from 6.34% through 3.10% to 2.02% (for  $x_1$ ) and from 11.57% through 5.38% to 3.44% (for  $x_2$ ), and meanwhile, the medians of the ten-step-ahead forecasting errors decreases from 6.43% through 2.49% to 1.38% (for  $x_1$ ) and from 6.49% through 2.54% to 1.42% (for  $x_2$ ).

Overall, combining the conclusions in parameter estimation Section 5.2, we conclude that integral matching consistently outperforms grey modelling in terms of parameter estimation, fitting and multi-step-ahead forecasting, as well as robustness to noise disturbance.

## 6. Real-world application

### 6.1. Data collection

Water supply plays an important role in a country's stability and development, while groundwater and surface water are two main water supply sources. Facing the global groundwater crisis [63], the Chinese government and enterprises have invested in many water projects over the past decades (including the waste-water treatment and reuse, rain collection, seawater desalinization, and others) in order to partially replace the groundwater losses that have occurred previously. For simplicity of presentation, the water supplies from these water projects are collectively referred to as the 'other water supply sources'. Forecasting these other water supplies accurately is clearly important in efforts to realize sustainable water resource development and research in the future. The total amount of other water supplies ( $10^9\text{m}^3$ ) is collected from the National Bureau of Statistics of China (<http://data.stats.gov.cn/english/easyquery.htm?cn=C01>). However, water supply records are incomplete and sparse. Only limited data from 2004 to 2018 are available, as shown in Table 3.

Using various estimated grey models, 5-step-ahead forecasts are generated in order to be consistent with the Five Year Plan of China, which sets the government's development goals for the next five-year stage. The data are divided into two parts: the data from 2004 to 2015 (12 samples taking 80% of the 15 samples) are used as a training dataset to construct the



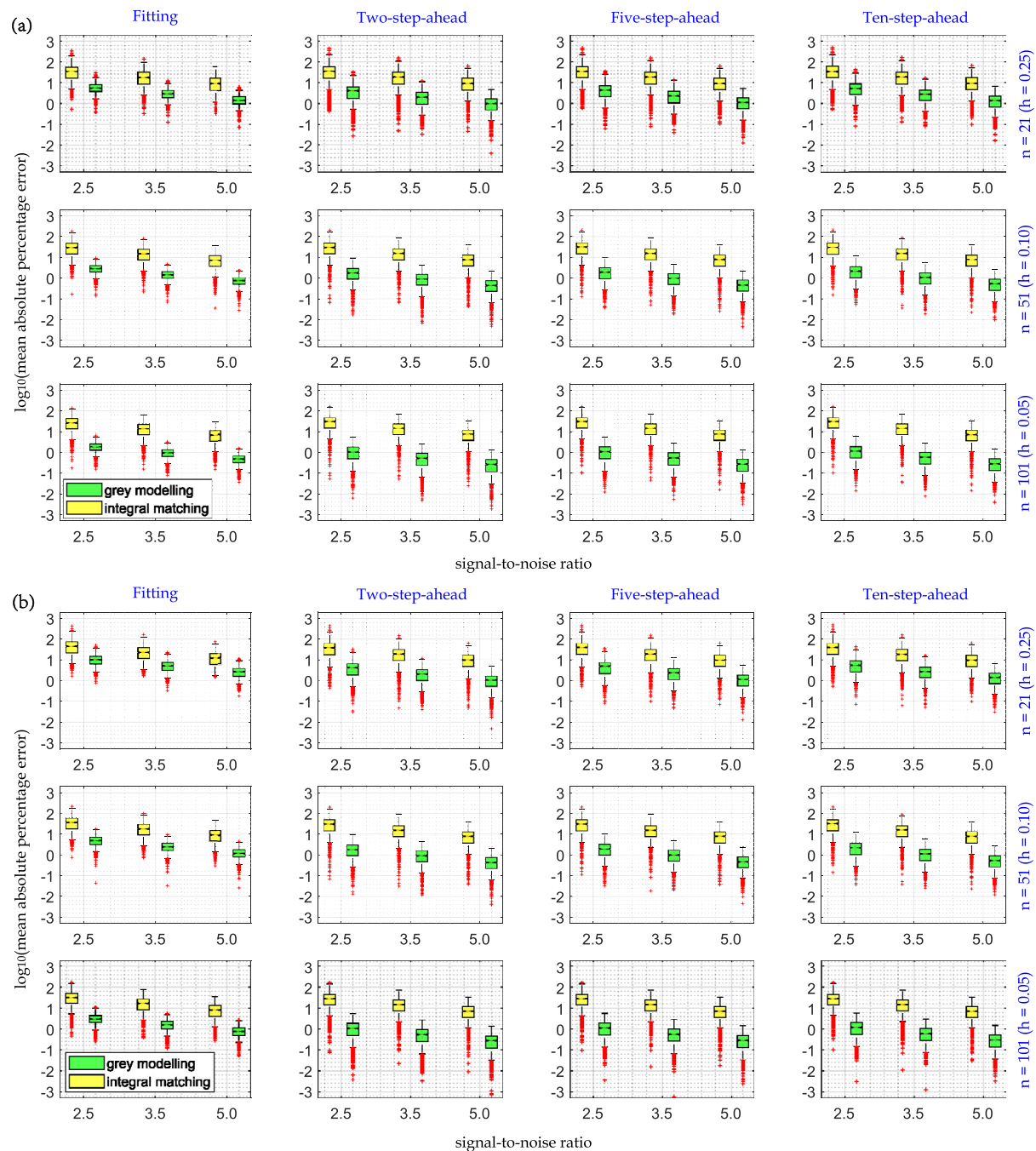


Fig. 3. Boxplots of the fitting and multi-step-ahead forecasting errors in 1000 replications: (a) the first component  $x_1(t)$ , (b) the second component  $x_2(t)$ .

Table 3  
The total amount of other water supplies in China from 2004 to 2018.

Year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Value	17.20	21.96	22.70	25.70	28.74	31.16	33.12	44.80	44.60	49.94	57.46	64.50	70.85	81.20	86.40



**Table 4**

Fitting and forecasting results obtained from integral matching-based and grey models.

Year	IMDE <sub>1</sub> <sup>i</sup>		IMDE <sub>2</sub> <sup>ii</sup>		IMDE <sub>3</sub> <sup>iii</sup>		IMDE <sub>4</sub> <sup>iv</sup>		IMDE <sub>5</sub> <sup>v</sup>		GPM(1,1,2) <sup>vi</sup>	
	Value	APE	Value	APE	Value	APE	Value	APE	Value	APE	Value	APE
2004	18.22	5.92	19.14	11.27	20.89	21.47	20.90	21.53	22.98	33.60	21.55	25.30
2005	20.43	6.99	21.12	3.84	21.66	1.38	21.67	1.33	22.32	1.66	21.48	2.17
2006	22.90	0.89	23.37	2.97	23.14	1.95	23.15	2.00	23.28	2.57	22.98	1.22
2007	25.68	0.09	25.95	0.97	25.32	1.49	25.33	1.45	25.47	0.90	25.16	2.10
2008	28.79	0.17	28.89	0.52	28.15	2.05	28.16	2.02	28.55	0.64	28.00	2.57
2009	32.28	3.59	32.24	3.47	31.61	1.45	31.62	1.47	32.29	3.63	31.47	0.99
2010	36.19	9.27	36.07	8.90	35.68	7.72	35.68	7.73	36.50	10.21	35.54	7.30
2011	40.58	9.43	40.43	9.74	40.31	10.02	40.32	10.01	41.10	8.26	40.18	10.31
2012	45.49	2.01	45.42	1.83	45.50	2.01	45.50	2.02	46.09	3.35	45.37	1.73
2013	51.01	2.14	51.10	2.33	51.20	2.53	51.21	2.54	51.60	3.32	51.08	2.29
2014	57.19	0.47	57.59	0.22	57.41	0.08	57.42	0.08	57.86	0.69	57.30	0.28
2015	64.12	0.58	64.99	0.76	64.10	0.62	64.10	0.62	65.23	1.14	63.99	0.79
MAPE <sub>in</sub> (%)		3.46		3.90		4.40		4.40		5.83		4.75
2016	71.90	1.47	73.43	3.65	71.24	0.55	71.24	0.55	74.25	4.80	71.14	0.40
2017	80.61	0.73	83.07	2.30	78.82	2.93	78.81	2.94	85.60	5.42	78.72	3.06
2018	90.38	4.61	94.07	8.87	86.81	0.48	86.80	0.46	100.15	15.91	86.72	0.37
MAPE <sub>out</sub> (%)		2.27		4.94		<b>1.32</b>		<b>1.32</b>		8.71		<b>1.28</b>
2019	101.33		106.61		95.21		95.18		118.98		95.12	
2020	113.61		120.93		103.98		103.93		143.40		103.89	

Note that the model structures are (i)  $\frac{d}{dt}x(t) = ax(t)$ , (ii)  $\frac{d}{dt}x(t) = ax(t) + c$ , (iii)  $\frac{d}{dt}x(t) = ax(t) + b_1t + c$ , (iv)  $\frac{d}{dt}x(t) = ax(t) + b_1t + b_2t^2 + c$ , (v)  $\frac{d}{dt}x(t) = ax(t) + b_1t + b_2t^2 + b_3t^3 + c$ , and (vi)  $\frac{d}{dt}y(t) = ay(t) + b_1t + b_2t^2 + c$ .

models and the data from 2016 to 2018 (3 samples taking 20% of the 15 samples) are used as test dataset to validate these identified models. In addition, the forecasts in 2019 and 2020 are used to evaluate the post-2018 behavior.

## 6.2. Results of integral matching-based and grey models

For simplicity of presentation, the integral matching-based differential equation model is abbreviated as IMDE. Table 4 shows the results of IMDE models and the only difference among these models lies in the forcing terms which are expressed by the linear combination of the time polynomial basis functions. The results show the following:

- In comparison with IMDE<sub>5</sub>, IMDE<sub>1</sub>–IMDE<sub>4</sub> have lower fitting errors (all the MAPE<sub>in</sub>s less than 5.0%) and forecasting errors (all the MAPE<sub>out</sub>s less than 5.0%), and meanwhile, IMDE<sub>5</sub> has more complex structure, indicating that IMDE<sub>1</sub>–IMDE<sub>4</sub> are superior to IMDE<sub>5</sub>.
- There are only small differences in the fitting error MAPE<sub>in</sub>s among IMDE<sub>1</sub>–IMDE<sub>4</sub>, whereas IMDE<sub>1</sub>, IMDE<sub>3</sub> and IMDE<sub>4</sub> have much lower forecasting errors in comparison with IMDE<sub>2</sub>, indicating that IMDE<sub>2</sub> is inferior to the other three.
- The fitting and forecasting errors of IMDE<sub>3</sub> and IMDE<sub>4</sub> are equal to each other, respectively, whereas IMDE<sub>4</sub> has an additional term  $b_2t^2$ , indicating that IMDE<sub>3</sub> is more parsimonious (parametrically efficient) and so better according to the Occam's razor principle [64]. In fact, the estimated parameters of IMDE<sub>4</sub> are  $\hat{a} = -0.0395$ ,  $\hat{c} = 0.4509$ ,  $\hat{b}_1 = 0.7717$ ,  $\hat{b}_2 = -0.0018$ ,  $\hat{\eta} = 20.9025$  where the estimates of  $\hat{b}_2$  is close to 0; and the other estimates are close to those of IMDE<sub>3</sub> with  $\hat{a} = -0.0458$ ,  $\hat{c} = 0.5761$ ,  $\hat{b}_1 = 0.7730$ ,  $\hat{\eta} = 20.8931$ .
- IMDE<sub>3</sub> has a higher fitting error MAPE<sub>in</sub> but lower forecasting error MAPE<sub>out</sub> than IMDE<sub>1</sub>. The reason for the former is that the estimated initial value of IMDE<sub>3</sub> is much greater than the true value. Without consideration of the first sample, the MAPE<sub>in</sub>s of IMDE<sub>1</sub> and IMDE<sub>3</sub> in the period from 2005 to 2015 are 3.24% and 2.85%, respectively, indicating that IMDE<sub>3</sub> is superior in terms of fitting and forecasting accuracies. In fact, the estimated parameters of IMDE<sub>1</sub> are  $\hat{a} = 0.1144$  and  $\hat{\eta} = 18.2176$ , resulting in a time response function expressed as  $x(t) = 16.2482 \exp(0.1144t)$ . This increases too fast to be used in medium- and long-term prediction (the APE = 4.61% in 2018 also tends to confirm this).

Overall, IMDE<sub>3</sub> is the best among the five IMDE models. Besides, corresponding to the model structure of IMDE<sub>3</sub>, we consider the second-order grey polynomial model [43], abbreviated as GPM(1,1,2) in Table 4, and analyse the similarities and differences, as shown in Table 4 and Fig. 4(a). Since the fitting values in 2004 are actually the estimated initial values, the differences between initial values are the main reason for the differences in performance. The details are present in the following:

The expressions of IMDE<sub>3</sub> and GPM(1,1,2) models are

$$\frac{d}{dt}x_m(t) = -0.04578x_m(t) + 0.7730t + 0.5761, \quad x(1) = 20.8931, \quad t \geq 1$$

and

$$\frac{d}{dt}y_g(t) = -0.04578y_g(t) + 0.3865t^2 + 0.9626t + 20.6123, \quad y(1) = 21.5509, \quad t \geq 1,$$

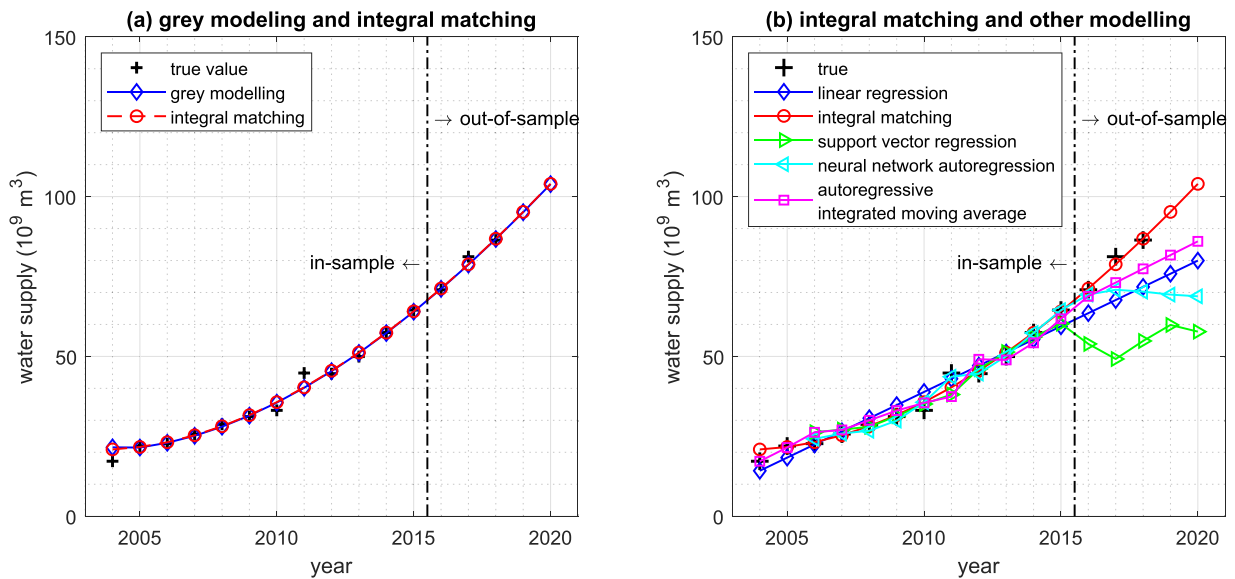


Fig. 4. Fitting and forecasting results of other water supply obtained from various models.

Table 5

Fitting and forecasting results obtained from other classical models.

Year	IMDE <sub>3</sub> <sup>i</sup>		LR <sup>ii</sup>		ARIMA <sup>iii</sup>		NNAR <sup>iv</sup>		SVR <sup>v</sup>	
	Value	APE	Value	APE	Value	APE	Value	APE	Value	APE
2004	20.89	21.47	14.22	17.30	17.19	0.07				
2005	21.66	1.38	18.33	16.51	21.50	2.09				
2006	23.14	1.95	22.44	1.13	26.26	15.68	24.21	6.66	26.26	15.70
2007	25.32	1.49	26.55	3.31	27.00	5.06	25.88	0.71	27.11	5.47
2008	28.15	2.05	30.66	6.68	30.00	4.38	26.74	6.96	27.98	2.64
2009	31.61	1.45	34.77	11.58	33.04	6.03	29.93	3.96	31.33	0.55
2010	35.68	7.72	38.88	17.38	35.46	7.07	35.63	7.57	35.11	6.00
2011	40.31	10.02	42.99	4.05	37.42	16.47	43.64	2.60	38.07	15.02
2012	45.50	2.01	47.10	5.59	49.10	10.09	44.63	0.07	46.02	3.18
2013	51.20	2.53	51.20	2.53	48.90	2.08	50.11	0.34	51.36	2.84
2014	57.41	0.08	55.31	3.74	54.24	5.60	57.32	0.24	56.05	2.46
2015	64.10	0.62	59.42	7.87	61.76	4.25	64.46	0.05	60.25	6.59
MAPE <sub>in</sub> (%)		4.40		8.14		6.57		2.92		6.05
2016	71.24	0.55	63.53	10.33	68.80	2.89	69.34	2.13	53.91	23.91
2017	78.82	2.93	67.64	16.70	73.10	9.98	70.85	12.74	49.19	39.42
2018	86.81	0.48	71.75	16.96	77.40	10.42	70.27	18.67	54.83	36.54
MAPE <sub>out</sub> (%)		<b>1.32</b>		14.66		7.76		11.18		33.29
2019	95.21		75.86		81.70		69.34		59.84	
2020	103.98		79.97		86.00		68.80		57.75	

Note that (i) IMDE<sub>3</sub> is the copy of that in Table 4 for easier comparison; (ii) the coefficient of determination is 0.9548; (iii) the model order is arima(0,1,0); (iv) the model type is feed-forward neural network with 2 lagged inputs and 3 nodes in the only hidden layer; (v) the model type is  $\epsilon$ -SVR with the embedding dimension equal to 2 and the kernel type being radial basis function.

respectively; then, the time response functions are

$$x_m(t) = 16.8847t + 377.1157 \exp(-0.04578t) - 356.2318, \quad t \geq 1$$

and

$$y_g(t) = 8.4424t^2 - 347.7895t - 8046.2287 \exp(-0.04578t) + 8047.0682, \quad t \geq 1,$$

which leads to the restored time response function expressed as

$$x_g(t) = \frac{y_g(t) - y_g(t-1)}{t - (t-1)} = 16.8847t + 376.9255 \exp(-0.04578t) - 356.2318, \quad t \geq 2.$$

By comparing  $x_m(t)$  with  $x_g(t)$ , IMDE<sub>3</sub> and GPM(1,1,2) models have the same time response functions except for the coefficients of the exponential terms which are caused by the estimated initial values. Overall, IMDE<sub>3</sub> has simpler modelling procedures and is easier to explain, so it is probably the best one to use in this application.

### 6.3. Comparison with other classical models

The integral matching-based model (IMDE<sub>3</sub>) is compared with the linear regression (LR), autoregressive integrated moving average (ARIMA), neural network autoregression (NNAR) and support vector regression (SVR). The calculation results associated with the absolute percentage errors (APEs) at each time instant are shown in Table 5 and Fig. 4(b). ARIMA and NNAR models are implemented by the `auto.arima` and `nnetar` functions in forecast package [65], and SVR model by the `svm` function in `e1071` package [66] in R software.

Table 5 shows that all the MAPE<sub>in</sub>s of the five models are less than 10% but their MAPE<sub>out</sub>s vary widely, indicating that they have almost the same fitting but different forecasting performance. The high coefficient of determination of LR indicates that the model order is appropriate, but it does not consider the autocorrelation in time series and thus results in poor forecasting performance. On the contrary, ARIMA consider the autocorrelation, thereby having a better performance than LR. NNAR obtains the smallest fitting error but a higher forecasting error, and SVR performs better in fitting but behaves poorly in forecasting. The reason for good fitting but bad forecasting performance of NNAR and SVR may be that the small modelling sample size leads to the over-fitting, although we have tried our best to avoid this by simplifying the model structures in the modelling process.

Overall, the integral matching-based model, IMDE<sub>3</sub>, has acceptable fitting error (MAPE<sub>in</sub> = 4.40%) and minimal forecasting error (MAPE<sub>out</sub> = 1.32%) and so can be considered the best in this example (see also Fig. 4(b)).

## 7. Conclusions

This paper has presented a unified framework for continuous-time grey models that subsume the single-variable, multi-variable, and multi-output grey models as special cases. It has shown that this unified form can be reduced into an equivalent differential equation and how the integral operator links the differential equations of unified and reduced models. It has also shown that structural parameters and initial condition of the reduced model can be simultaneously estimated by the integral matching approach. We see, therefore, that grey and reduced models are both concerned with the data-based modelling in terms of differential equations. The former uses the Cusum operator explicitly, while the latter implicitly employs the Cusum operator via a numerical discretization scheme of the integral operator. As pointed out in Section 4.2, in practical applications, we recommend the integral matching-based reduction form due to its advantages over grey models, including higher accuracy, simpler modelling procedures and easier explanation of modelling results.

This paper has not described the impact of observed errors on the estimates of grey models and reduced models. The presence of errors-in-variables (also known as measurement error in statistics) induces an asymptotic bias on the parameter estimates which is a function of the signal-to-noise ratio on  $\mathbf{x}(t_k)$  and is zero only when there is no noise. Other methods of estimation have advantages in this regard, such as the refined instrumental variable estimation; another possibility is two-stage linear least squares where the original time-series observations are de-noised at the first stage and then the parameters are estimated using linear least squares at the second stage.

Besides, both grey and reduced models use linear least squares-based statistical estimation but do not consider the statistical properties of their corresponding estimators. Consequently, future research should be focused on the statistical properties concerning minimum variance and asymptotic analysis, that is, the asymptotic consistency and efficiency respectively measuring the convergence and variance of the estimators as the time interval  $h$  tends to 0.

### Data availability

Data and codes could be seen in the following link: <https://github.com/weibl9/CTGMs>.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

This work was supported by the National Natural Science Foundation of China (72171116, 71671090) and the Fundamental Research Funds for the Central Universities of China (NP2020022). This research was completed during the first author's visit at Lancaster University; thanks to Wlodek Tych for hosting my visit. The authors are most grateful to Peter C. Young for his valuable suggestions. This does not mean that he necessarily shares the views stated in the paper, and the authors, of course, responsible for any errors or omissions. Also thanks to the associate editor and referee for useful suggestions that have improved the quality of this paper.

## References

- [1] J. Deng, Control problems of grey systems, *Syst. Control Lett.* 1 (5) (1982) 288–294.
- [2] J. Deng, Grey theory and methods in social and economic system (in Chinese), *Soc. Sci. China* (6) (1984) 47–60.
- [3] S. Liu, Y. Yang, J. Forrest, *Grey Data Analysis: Methods, Models and Applications*, Springer-Verlag, Singapore, 2017.
- [4] N. Xie, R. Wang, A historic review of grey forecasting models, *J. Grey Syst.* 29 (4) (2017) 1–29.
- [5] D.-C. Li, C.-J. Chang, W.-C. Chen, C.-C. Chen, An extended grey forecasting model for omnidirectional forecasting considering data gap difference, *Appl. Math. Model.* 35 (10) (2011) 5051–5058.
- [6] J. Nešetřil, J. Matoušek, *Invitation to Discrete Mathematics*, 2nd, Oxford University Press, 2009.
- [7] X. Xiao, H. Guo, S. Mao, The modeling mechanism, extension and optimization of grey GM(1,1) model, *Appl. Math. Model.* 38 (5–6) (2014) 1896–1910.
- [8] X. Yang, Z. Fang, Y. Yang, D. Mba, X. Li, A novel multi-information fusion grey model and its application in wear trend prediction of wind turbines, *Appl. Math. Model.* 71 (2019) 543–557.
- [9] R.K. Colwell, C. Mao, J. Chang, Interpolating, extrapolating, and comparing incidence-based species accumulation curves, *Ecology* 85 (10) (2004) 2717–2727.
- [10] S. Chatterjee, P. Qiu, Distribution-free cumulative sum control charts using bootstrap-based control limits, *Ann. Appl. Stat.* 3 (1) (2009) 349–369.
- [11] Z. Zhao, J. Wang, J. Zhao, Z. Su, Using a grey model optimized by differential evolution algorithm to forecast the per capita annual net income of rural households in china, *Omega (Westport)* 40 (5) (2012) 525–532.
- [12] D.-C. Li, C.-W. Yeh, C.J. Chang, An improved grey-based approach for early manufacturing data forecasting, *Comput. Ind. Eng.* 57 (4) (2009) 1161–1167.
- [13] D.-C. Li, C.-J. Chang, C.-C. Chen, W.C. Chen, Forecasting short-term electricity consumption using the adaptive grey-based approach—an Asian case, *Omega (Westport)* 40 (6) (2012) 767–773.
- [14] C.-J. Chang, D.-C. Li, Y.-H. Huang, C.-C. Chen, A novel gray forecasting model based on the box plot for small manufacturing data sets, *Appl. Math. Comput.* 265 (2015) 400–408.
- [15] C.-S. Shih, Y.-T. Hsu, J. Yeh, P.-C. Lee, Grey number prediction using the grey modification model with progression technique, *Appl. Math. Model.* 35 (3) (2011) 1314–1321.
- [16] Y. Dang, S. Liu, K. Chen, The GM models that  $x(n)$  be taken as initial value, *Kybernetes* 33 (2) (2004) 247–254.
- [17] J. Xu, T. Tan, M. Tu, L. Qi, Improvement of grey models by least squares, *Expert Syst. Appl.* 38 (11) (2011) 13961–13966.
- [18] Y. Wang, Q. Liu, J. Tang, W. Cao, X. Li, Optimization approach of background value and initial item for improving prediction precision of GM(1,1) model, *J. Syst. Eng. Electron.* 25 (1) (2014) 77–82.
- [19] C.I. Chen, S.J. Huang, The necessary and sufficient condition for GM(1,1) grey prediction model, *Appl. Math. Comput.* 219 (11) (2013) 6152–6162.
- [20] J. Liu, X.P. Xiao, J.H. Guo, S.H. Mao, Error and its upper bound estimation between the solutions of GM(1,1) grey forecasting models, *Appl. Math. Comput.* 246 (C) (2014) 648–660.
- [21] T.L. Tien, A new grey prediction model FGM(1,1), *Math. Comput. Model.* 49 (7) (2009) 1416–1426.
- [22] T. Yao, S. Liu, N. Xie, On the properties of small sample of GM(1,1) model, *Appl. Math. Model.* 33 (4) (2009) 1894–1903.
- [23] L. Wu, S. Liu, L. Yao, S. Yan, The effect of sample size on the grey system model, *Appl. Math. Model.* 37 (9) (2013) 6577–6583.
- [24] R. Guo, D. Guo, Random fuzzy variable foundation for grey differential equation modeling, *Soft comput.* 13 (2) (2009) 185–201.
- [25] J. Cui, S. Liu, B. Zeng, N. Xie, A novel grey forecasting model and its optimization, *Appl. Math. Model.* 37 (6) (2013) 4399–4406.
- [26] P.-Y. Chen, H.-M. Yu, Foundation settlement prediction based on a novel NGM model, *Math. Problems Eng.* 2014 (2014) 242809.
- [27] S. Ding, R. Li, S. Wu, W. Zhou, Application of a novel structure-adaptive grey model with adjustable time power item for nuclear energy consumption forecasting, *Appl. Energy* 298 (2021) 117114.
- [28] D. Luo, B. Wei, Grey forecasting model with polynomial term and its optimization, *J. Grey Syst.* 29 (3) (2017) 58–69.
- [29] X. Ma, Y. Hu, Z. Liu, A novel kernel regularized nonhomogeneous grey model and its applications, *Commun. Nonlinear Sci. Numer. Simul.* 48 (2016) 51–62.
- [30] M. Evans, An alternative approach to estimating the parameters of a generalised Grey Verhulst model: an application to steel intensity of use in the UK, *Expert Syst. Appl.* 41 (4) (2014) 1236–1244.
- [31] C.-I. Chen, H. Chen, S.-P. Chen, Forecasting of foreign exchange rates of Taiwan's major trading partners by novel nonlinear Grey Bernoulli model NGBM(1,1), *Commun. Nonlinear Sci. Numer. Simul.* 13 (6) (2008) 1194–1204.
- [32] L. Yang, N. Xie, Integral matching-based nonlinear grey Bernoulli model for forecasting the coal consumption in China, *Soft Comput.* 25 (7) (2021) 5209–5223.
- [33] T.-L. Tien, A research on the grey prediction model GM(1,n), *Appl. Math. Comput.* 218 (9) (2012) 4903–4916.
- [34] B. Zeng, C. Luo, S. Liu, Y. Bai, C. Li, Development of an optimization method for the GM(1,N) model, *Eng. Appl. Artif. Intell.* 55 (2016) 353–362.
- [35] X. Ma, Z.-b. Liu, The kernel-based nonlinear multivariate grey model, *Appl. Math. Model.* 56 (2018) 217–238.
- [36] Z.-X. Wang, D.-J. Ye, Forecasting chinese carbon emissions from fossil energy consumption using non-linear grey multivariable models, *J. Clean. Prod.* 142 (2017) 600–612.
- [37] L. Yu, X. Ma, W. Wu, Y. Wang, B. Zeng, A novel elastic net-based NGBMC(1,n) model with multi-objective optimization for nonlinear time series forecasting, *Commun. Nonlinear Sci. Numer. Simul.* 96 (2021) 105696.
- [38] P. Xiong, Y. Zhang, B. Zeng, T. Yao, MGM(1,m) model based on interval grey number sequence and its applications, *Grey Systems: Theory and Application* 7 (3) (2017) 310–319.
- [39] X. Guo, S. Liu, L. Wu, Y. Gao, Y. Yang, A multi-variable grey model with a self-memory component and its application on engineering prediction, *Eng. Appl. Artif. Intell.* 42 (2015) 82–93.
- [40] X. Guo, S. Liu, Y. Yang, A prediction method for plasma concentration by using a nonlinear grey Bernoulli combined model based on a self-memory algorithm, *Comput. Biol. Med.* 105 (2019) 81–91.
- [41] L. Wu, S. Liu, Y. Wang, Grey Lotka–Volterra model and its application, *Technol. Forecast. Soc. Change* 79 (9) (2012) 1720–1730.
- [42] H.-T. Wang, T.-C. Wang, Application of the grey Lotka–Volterra model to forecast the diffusion and competition analysis of the tv and smartphone industries, *Technol. Forecast. Soc. Change* 106 (2016) 37–44.
- [43] B. Wei, N. Xie, A. Hu, Optimal solution for novel grey polynomial prediction model, *Appl. Math. Model.* 62 (2018) 717–727.
- [44] S. Ding, R. Li, S. Wu, A novel composite forecasting framework by adaptive data preprocessing and optimized nonlinear grey Bernoulli model for new energy vehicles sales, *Commun. Nonlinear Sci. Numer. Simul.* 99 (2021) 105847.
- [45] B. Zeng, C. Li, Improved multi-variable grey forecasting model with a dynamic background-value coefficient and its application, *Comput. Ind. Eng.* 118 (2018) 278–290.
- [46] G. Comert, N. Begashaw, N. Huynh, Improved grey system models for predicting traffic parameters, *Expert Syst. Appl.* 177 (2021) 114972.
- [47] Y.-H. Lin, P.-C. Lee, Novel high-precision grey forecasting model, *Autom. Constr.* 16 (6) (2007) 771–777.
- [48] Y.-H. Lin, P.-C. Lee, T.-P. Chang, Adaptive and high-precision grey forecasting model, *Expert Syst. Appl.* 36 (6) (2009) 9658–9662.
- [49] E. Kayacan, B. Ulutas, O. Kaynak, Grey system theory-based models in time series prediction, *Expert Syst. Appl.* 37 (2) (2010) 1784–1789.
- [50] P. Xiong, Y. Dang, T. Yao, Z. Wang, Optimal modeling and forecasting of the energy consumption and production in China, *Energy* 77 (2014) 623–634.
- [51] L.-M. Kung, S.-W. Yu, Prediction of index futures returns and the analysis of financial spillovers—a comparison between GARCH and the grey theorem, *Eur. J. Oper. Res.* 186 (3) (2008) 1184–1200.
- [52] P. Xiong, S. Huang, M. Peng, X. Wu, Examination and prediction of fog and haze pollution using a multi-variable grey model based on interval number sequences, *Appl. Math. Model.* 77 (2020) 1531–1544.

- [53] P. Xiong, Y. Dang, X. Wu, X. Li, Combined model based on optimized multi-variable grey model and multiple linear regression, *J. Syst. Eng. Electron.* 22 (4) (2011) 615–620.
- [54] I. Dattner, C.A. Klaassen, Optimal rate of direct estimators in systems of ordinary differential equations linear in functions of the parameters, *Electron. J. Stat.* 9 (2) (2015) 1939–1973.
- [55] I. Dattner, A model-based initial guess for estimating parameters in systems of ordinary differential equations, *Biometrics* 71 (4) (2015) 1176–1184.
- [56] B. Wei, N. Xie, L. Yang, Understanding cumulative sum operator in grey prediction model with integral matching, *Commun. Nonlinear Sci. Numer. Simul.* 82 (2020) 105076.
- [57] S. Ding, R. Li, Forecasting the sales and stock of electric vehicles using a novel self-adaptive optimized grey model, *Eng. Appl. Artif. Intell.* 100 (2021) 104148.
- [58] C.F. Van Loan, G.H. Golub, *Matrix Computations*, 4th, Johns Hopkins University Press, 2013.
- [59] B. Wei, N. Xie, Y. Yang, Data-based structure selection for unified discrete grey prediction model, *Expert Syst. Appl.* 136 (2019) 264–275.
- [60] P.C. Young, D.J. Pedregal, W. Tych, Dynamic harmonic regression, *J. Forecast* 18 (6) (1999) 369–394.
- [61] W. Tych, D.J. Pedregal, P.C. Young, J. Davies, An unobserved component model for multi-rate forecasting of telephone call demand: the design of a forecasting support system, *Int. J. Forecast* 18 (4) (2002) 673–695.
- [62] P.C. Young, Data-based mechanistic modelling and forecasting globally averaged surface temperature, *Int. J. Forecast.* 34 (2018) 315–334.
- [63] J.S. Famiglietti, The global groundwater crisis, *Nat. Clim. Chang.* 4 (11) (2014) 945.
- [64] P. Young, S. Parkinson, M. Lees, Simplicity out of complexity in environmental modelling: Occam's razor revisited, *J. Appl. Stat.* 23 (2–3) (1996) 165–210.
- [65] R.J. Hyndman, Y. Khandakar, Automatic time series forecasting: the forecast package for R, *J. Stat. Softw.* 27 (3) (2008) 1548–7660.
- [66] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, F. Leisch, e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien, 2021. R package version 1.7–7.