Highlights

- A novel framework is proposed to unify the existing single-variable, multi-variable and multi-output grey models.
- The unified model is simplified into a reduced-order ordinary differential equation model by introducing integral transformation.
- The cumulative sum operator proves to be the discrete approximation of the integral transformation.
- The integral matching is proposed to estimate the structural parameter and initial value simultaneously.
- Large-scale simulations are conducted to evaluate the finite-sample performance under different sample size and signal-to-noise ratio configurations.

On novel framework for continuous-time grey models: an integral matching perspective

Baolei Wei[a], Naiming Xie[a,*]

[a] *College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, PR China*

Abstract

Since most of the researches about grey forecasting models are 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 models is proposed and simplified into a reduced-order ordinary differential equation. Then, the integral matching that consists of integral transformation 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 transformation formula. 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 simulation studies, and a real-world example shows how to apply and further verify the reconstructed model.

Keywords: grey system theory, grey forecasting model, cumulative sum operator, ordinary differential equation, integral matching

1. Introduction

Time series forecasting has been a challenging and important topic in methodology and practical applications ranging from natural science through engineering technique to social science. There exist a variety of time series forecasting methods including the popular statistical regression and the emerging machine learning, such as exponential smoothing [1], Box-Jenkins discrete-time [2] and continuous-time [3] models, support vector regression [4], and neural network [5], etc. Each model has its own advantages and limitation; see [6] for a review.

By taking time series as the behaviour records of the underlying dynamics, the time series forecasting problem is converted to that of forecasting the evolution of a dynamic system. The key is to discover the underlying dynamics from time-series observations, and the relevant approaches can be referred to as data-based/data-driven dynamical modelling (see [7–9] and references therein). In a generalized sense, the grey forecasting model proposed by Deng [10] in 1984, also uses this kind of approach. In contrast to other methods, however, grey forecasting models are constructed based on the prior assumption that the cumulation and release in many generalized energy systems conform to an exponential law [11]. Grey forecasting models utilize the cumulative sum (Cusum) operator to mine the exponential characteristic hidden in the original time series and then the continuous-time dynamic (ordinary differential equation) to fit the Cusum series. Generally, grey models are usually termed as GM(φ, ψ) where φ denotes the order of derivative and ϕ denotes the number of variables. Over the past three decades, research has shown consistently that grey models are promising in solving a class of time series prediction problems, especially in the case of small-sample data sets [12].

*Corresponding author.

Email addresses: weibl@nuaa.edu.cn (Baolei Wei), xienaiming@nuaa.edu.cn (Naiming Xie)

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

1.1. The cumulative sum operator and its inverse

Definition 1. *For the time series* $X(t) = \{x(t_1), x(t_2), \cdots, x(t_n)\}$*, the Cusum series is defined as* $Y(t) = \{y(t_1), y(t_2), \cdots, y(t_n)\}$ *where* $y(t_k) = \sum_{i=1}^{k} h_i x(t_i)$*,* $h_1 = 1$ *and* $h_k = t_k - t_{k-1}$*,* $k \geq 2$*. Corresponding, the inverse Cusum operator is defined as* $x(t_1) = \frac{1}{h_1} y(t_1)$ *and* $x(t_k) = \frac{1}{h_k}\left(y(t_k) - y(t_{k-1})\right)$ *when* $k \geq 2$*.*

Note that the first time interval h_1 is set as 1 according to the empty product principle in mathematics [13]. 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 [14–16].

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 modelling results [10]. In these terms, the Cusum operator can be viewed as a data preprocessing technique which is used to mine the pattern hidden in the original time series.

There are some advantages of Cusum operator, especially from a data visualization perspective. Considering the following example, it is difficult to identify the hidden pattern from the line graph of the original time series (see Figure 1(a)), while the quasi-exponential characteristic is obvious in the cumulative sum series (see Figure 1(b)).

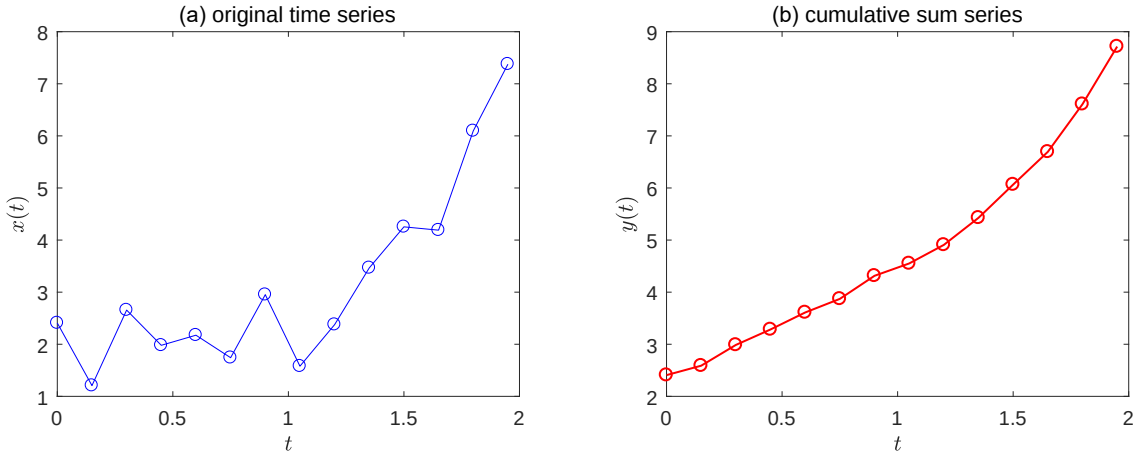

Figure 1: Plots of the original time series and the corresponding cumulative sum series.

In fact, a grey forecasting model is not the only one that uses the Cusum operator to mine the hidden patterns. The similar idea was also employed in other disciplines, such as ecology and statistical process control [17–19].

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

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

and the discrete-time difference equation

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

where $k = 2, 3, \cdots, n$, $n \geq 4$, 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 as

$$\begin{bmatrix} \hat{a} & \hat{b} \end{bmatrix}^{\mathsf{T}} = \left(\boldsymbol{D}^{\mathsf{T}}\boldsymbol{D}\right)^{-1}\boldsymbol{D}^{\mathsf{T}}\boldsymbol{y} \tag{3}$$

where

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

By setting the initial value parameter

$$y(t_1) = \eta, \tag{4}$$

the time response function is calculated as

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

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 $\hat{x}(t_k) = \frac{1}{h_k}\left[\hat{y}(t_k) - \hat{y}(t_{k-1})\right]$, $k = 2, 3, \cdots, n+r$, 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) Equation (2) is the numerical discretization-based approximation of equation (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. For example, Zhao et al. [20] used the differential evolution algorithm to search the optimal background coefficient. On the other hand, by using the trapezoid formula in each closed interval $[t_{k-1}, t_k]$, the discrete-time equation (2) is modified to be the following form

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

where $\lambda_k \in [0, 1]$, $k = 2, 3, \cdots, n$. In each closed interval $[t_{k-1}, t_k]$, the background coefficient λ_k is adaptively computed by using the triangular membership function rule [21–23] or the exponential function rule [24, 25].

(ii) The initial value parameter was set as $\eta = x(t_1)$ in the classical research [11]. Therefore, based on the intuition that new observations are likely to be more informative than historical ones, Dang et al. [26] proposed the initial value selection strategy $\xi = x(t_n)$. On the other hand, in order to further reduce the fitting error, Wang et al. [27] and Xu et al. [28] proposed the least-square strategy:

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

where $\hat{y}(t_k)$ is calculated according to the time response equation (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, Wang et al. [24] proposed a combined method which consists of an explicit formula for background coefficient calculation and the least-square strategy for initial value selection.

In addition, another research focus of GM(1,1) model is the property analysis, such as the necessary and sufficient condition for modelling [29], the error bound estimation among the variants of GM(1,1) model [30], the influence of data transformation on modelling accuracy [15, 31] and the effect of sample size on modelling performance [32, 33].

1.3. The extended grey forecasting models

Since the procedures of GM(1,1) provide a clear paradigm for grey modelling, Guo and Guo [34] probed into the mathematical principles for model extension 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: extending the linear model into the nonlinear ones and also 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 extended 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: 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)	$\frac{d}{dt}y(t) = ay(t) + b$	$x(k) = a\frac{y(k-1)+y(k)}{2} + b$	[11]
	NGM(1,1,k)	$\frac{d}{dt}y(t) = ay(t) + bt$	$x(k) = a\frac{y(k-1)+y(k)}{2} + bk$	[35]
	NGM(1,1,k,c)	$\frac{d}{dt}y(t) = ay(t) + b_1 t + b_0$	$x(k) = a\frac{y(k-1)+y(k)}{2} + b_1 k + b_0$	[36]
	GM(1,1,t^α)	$\frac{d}{dt}y(t) = ay(t) + b_1 t^\alpha + b_0$	$x(k) = a\frac{y(k-1)+y(k)}{2} + b_1 k^\alpha + b_0$	[37]
	GPM(1,1,N)	$\frac{d}{dt}y(t) = ay(t) + \sum_{j=0}^{N} b_j t^j$	$x(k) = a\frac{y(k-1)+y(k)}{2} + \sum_{j=0}^{N} b_j \frac{k^{j+1}-(k-1)^j}{j+1}$	[38]
	KRNGM(1,1)	$\frac{d}{dt}y(t) = ay(t) + \sum_{j=0}^{N} b_j \phi_j(t)$	$x(k) = a\frac{y(k-1)+y(k)}{2} + \sum_{j=0}^{N} b_j \frac{\phi_j(k)+\phi_j(k-1)}{2}$	[39]
Nonlinear	Verhulst	$\frac{d}{dt}y(t) = ay(t) + b[y(t)]^2$	$x(k) = a\frac{y(k-1)+y(k)}{2} + b\left(\frac{y(k-1)+y(k)}{2}\right)^2$	[40]
	NBGM(1,1)	$\frac{d}{dt}y(t) = ay(t) + b[y(t)]^\gamma$	$x(k) = a\frac{y(k-1)+y(k)}{2} + b\left(\frac{y(k-1)+y(k)}{2}\right)^\gamma$	[41]

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

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.

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) [11]; then on this basis Tien [42] came to the complete extension by introducing a control parameter. Subsequently, researchers proposed some other multi-variable models including the NGM(1,N) [43] which is the extension of NGM(1,1,k,c), the KGM(1,N) [44] which is the extension of KRNGM(1,1), and the NBGM(1,N) [45] 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 [46] associated with its improvement coupled with self-memory [47], and the nonlinear MNBGM(1,N) model coupled with self-memory [48], as well as the nonlinear grey Lotka–Volterra model [49, 50].

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, including the GPM(1,1,N) model [51], the Verhulst model [52] and the NBGM(1,1) model [53–56], (ii) the multi-variable cases, including the GM(1,N) model [57] and the NGM(1,N) model [58].

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 the paper are as follows:

(i) 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.

(ii) 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.

(iii) 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:

(i) 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.

(ii) By introducing an integral transformation, which can be viewed as the continuous generalization of the original time series, the Cusum operator is proved to be the Euler's formula-based numerical discretization of the continuous integral transformation; and, subsequently, the mechanism of the Cusum operator is further explained within the parameter estimation process.

(iii) 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 transformation to explain the Cusum mechanism, obtain the equivalent reduced-order differential equation, and present integral matching as a method for simultaneously estimating the model structure 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 procedure of the grey model using that of the integral matching-based one; section 5 provides large-scale simulation studies that compare the two approaches from a statistical significance viewpoint; 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. All computational results in this work are reproducible; data and codes are available at: `https://github.com/weibl9/CTGMs/archive/master.zip`.

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 $\boldsymbol{x}(t) \in \mathbb{R}^d$, the observational data sampled at time instants $t_1, t_2, \cdots, t_n$ are arranged into the following matrix form:

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

and subsequently the corresponding cumulative sum form is expressed as

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

where the Cusum column vector is calculated according to the formula $\boldsymbol{y}(t_k) = \sum_{i=1}^{k} h_i \boldsymbol{x}(t_i)$ 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

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

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

First, numerical discretization-based gradient matching is employed to estimate the structural parameters in equation (6). This method can be divided into the following two steps.

In the first step, using the trapezoid formula gives the discrete-time equation

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

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

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

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

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

where

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

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

$$\mathcal{L}(\boldsymbol{\Xi}) = \|\boldsymbol{X} - \boldsymbol{\Theta}(\boldsymbol{y}, \boldsymbol{u})\boldsymbol{\Xi}\|_{\mathsf{F}}^2 \tag{10}$$

where $\|\cdot\|_{\mathsf{F}}$ 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

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

and the subsequent least-square estimate

$$\hat{\boldsymbol{\Xi}} = \left(\boldsymbol{\Theta}^{\mathsf{T}}(\boldsymbol{y},\boldsymbol{u})\boldsymbol{\Theta}(\boldsymbol{y},\boldsymbol{u})\right)^{-1}\boldsymbol{\Theta}^{\mathsf{T}}(\boldsymbol{y},\boldsymbol{u})\boldsymbol{X}. \tag{11}$$

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

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

where $\boldsymbol{y}(t_1)=\boldsymbol{\eta}\in\mathbb{R}^d$ 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 ($\boldsymbol{\eta}=\boldsymbol{x}(t_1)$) which results in the time response function

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

(ii) the fixed last point strategy ($\boldsymbol{\eta}=\boldsymbol{x}(t_n)$) which results in the time response function

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

(iii) the least-squares strategy, in which $\boldsymbol{\eta}$ is obtained by minimizing the least-square objective function $\mathcal{L}(\boldsymbol{\eta}) = \sum_{k=1}^{n}\|\boldsymbol{y}(t_k)-\bar{\boldsymbol{y}}(t_k)\|_2^2$ and subsequently, the time response function is

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

Finally, substituting the time points $\{t_k\}_{k=1}^{n+r}$ into equation (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:

$$\hat{\boldsymbol{x}}(t_k) = \begin{cases}\hat{\boldsymbol{\eta}}, & k=1\\ \frac{\hat{\boldsymbol{y}}(t_k)-\hat{\boldsymbol{y}}(t_{k-1})}{t_k-t_{k-1}}, & k=2,3,\cdots,n+r\end{cases}. \tag{16}$$

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 $\boldsymbol{u}(t)=\left[u_\iota(t)\right]_{p\times 1}$ 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 $u_\iota(t)=t^\iota$, $\iota=1,2,\cdots,p$. By inserting $u_\iota(t)$ back into equations (6) and (8), we obtain the continuous- and discrete-time equations respectively expressed as

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

and

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

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

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

and

$$x(t_k)=a\frac{y(t_{k-1})+y(t_k)}{2}+\sum_{\iota=1}^{p/2}b_{2\iota-1}\frac{\sin(\iota t_{k-1})+\sin(\iota t_k)}{2}+\sum_{\iota=1}^{p/2}b_{2\iota}\frac{\cos(\iota t_{k-1})+\cos(\iota t_k)}{2}+c+\varepsilon(t_k).$$

Remark 1 shows that this unified model subsumes a number of special families of grey models, although here only two of them are considered. 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 covers the hybrid grey models which uses the Fourier series fitting to modify the residual errors of GM(1,1) model [59–62].

Remark 2. *Let the dimension of the state vector be* $d=1$. *If the vector* $\boldsymbol{u}(t)=\left[u_\iota(t)\right]_{p\times1}$ *consists of other state variables, i.e.,* $u_\iota(t)=y_\iota(t)$, $\iota=1,2,\cdots,p$, *then the unified model yields the multi-variable grey forecasting model [42, 63], whose continuous- and discrete-time equations are*

$$\frac{d}{dt}y(t)=ay(t)+\sum_{\iota=1}^{p}b_\iota y_\iota(t)+c$$

and

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

If the vector $\boldsymbol{u}(t)=\left[u_\iota(t)\right]_{p\times1}$ consists of a mixture of other state variables and known fixed basis functions of time, then combining the extension principle in Remark 1 gives

$$\frac{d}{dt}y(t)=ay(t)+\sum_{\iota=1}^{p_1}b_\iota y_\iota(t)+\sum_{\iota=1}^{p_2}b_\iota u_\iota(t)+c$$

and

$$x(t_k)=a\frac{y(t_{k-1})+y(t_k)}{2}+\sum_{\iota=1}^{p_1}b_\iota\frac{y_\iota(t_{k-1})+y_\iota(t_k)}{2}+\sum_{\iota=1}^{p_2}b_\iota\frac{u_\iota(t_{k-1})+u_\iota(t_k)}{2}+c+\varepsilon(t_k),$$

which shows that this model yields many multi-variable grey forecasting models, such as the aforementioned NGM(1,N) model [43] ($p_1=N-1$, $p_2=1$ 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* $d\geq2$. *If the vector* $\boldsymbol{u}(t)=\left[u_\iota(t)\right]_{p\times1}$ *is equal to* $\mathbf{0}$, *i.e.,* $u_\iota(t)=0$, $\iota=1,2,\cdots,p$, *then the unified model yields the multi-output grey forecasting model [46, 64], whose continuous- and discrete-time equations are*

$$\begin{cases}\frac{d}{dt}y_1(t)=\sum_{\iota=1}^{d}a_{1,\iota}y_\iota(t)+c_1\\ \frac{d}{dt}y_2(t)=\sum_{\iota=1}^{d}a_{2,\iota}y_\iota(t)+c_2\\ \quad\vdots\\ \frac{d}{dt}y_d(t)=\sum_{\iota=1}^{d}a_{d,\iota}y_\iota(t)+c_d\end{cases}\quad\textit{and}\quad\begin{cases}x_1(t_k)=\sum_{\iota=1}^{d}a_{1,\iota}\frac{y_\iota(t_{k-1})+y_\iota(t_k)}{2}+c_1\\ x_2(t_k)=\sum_{\iota=1}^{d}a_{2,\iota}\frac{y_\iota(t_{k-1})+y_\iota(t_k)}{2}+c_2\\ \quad\vdots\\ x_d(t_k)=\sum_{\iota=1}^{d}a_{d,\iota}\frac{y_\iota(t_{k-1})+y_\iota(t_k)}{2}+c_d\end{cases}.$$

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

$$\begin{cases}\frac{d}{dt}y_1(t)=\sum\limits_{\iota=1}^{d}a_{1,\iota}y_\iota(t)+\sum\limits_{\iota=1}^{p}b_{1,\iota}u_\iota(t)+c_1\\ \frac{d}{dt}y_2(t)=\sum\limits_{\iota=1}^{d}a_{2,\iota}y_\iota(t)+\sum\limits_{\iota=1}^{p}b_{2,\iota}u_\iota(t)+c_2\\ \quad\vdots\\ \frac{d}{dt}y_d(t)=\sum\limits_{\iota=1}^{d}a_{d,\iota}y_\iota(t)+\sum\limits_{\iota=1}^{p}b_{d,\iota}u_\iota(t)+c_d\end{cases}$$

and

$$\begin{cases}x_1(t_k)=\sum\limits_{\iota=1}^{d}a_{1,\iota}\frac{y_\iota(t_{k-1})+y_\iota(t_k)}{2}+\sum\limits_{\iota=1}^{p}b_{1,\iota}\frac{u_\iota(t_{k-1})+u_\iota(t_k)}{2}+c_1\\ x_2(t_k)=\sum\limits_{\iota=1}^{d}a_{2,\iota}\frac{y_\iota(t_{k-1})+y_\iota(t_k)}{2}+\sum\limits_{\iota=1}^{p}b_{2,\iota}\frac{u_\iota(t_{k-1})+u_\iota(t_k)}{2}+c_2\\ \quad\vdots\\ x_d(t_k)=\sum\limits_{\iota=1}^{d}a_{d,\iota}\frac{y_\iota(t_{k-1})+y_\iota(t_k)}{2}+\sum\limits_{\iota=1}^{p}b_{d,\iota}\frac{u_\iota(t_{k-1})+u_\iota(t_k)}{2}+c_d\end{cases},$$

which can not only yield the existing multi-output models, such as the ones coupling self-memory theory [47] and multiple regression [65], but suggest other novel ones.

Remarks 1–3 show that the unified equation (6) has the capacity to represent the continuous-time grey single-variable, multi-variable and multi-output 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 transformation and least squares [66–68], 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* $\boldsymbol{y}_{\rm tr}(t_k)=\boldsymbol{y}(t_k)+\boldsymbol{\xi}$, $k=1,2,\cdots,n$, *where* $\boldsymbol{\xi}\in\mathbb{R}^d$ *is a real vector. Then (i) the estimated parameters satisfy* $\hat{\boldsymbol{A}}_{\rm tr}=\hat{\boldsymbol{A}}$, $\hat{\boldsymbol{B}}_{\rm tr}=\hat{\boldsymbol{B}}$, $\hat{\boldsymbol{c}}_{\rm tr}=\hat{\boldsymbol{c}}-\hat{\boldsymbol{A}}\boldsymbol{\xi}$; *(ii) the time response functions satisfy* $\hat{\boldsymbol{y}}_{\rm tr}(t)=\hat{\boldsymbol{y}}(t)+\boldsymbol{\xi}$; *(iii) the fitting and forecasting values of the original time series satisfy* $\hat{\boldsymbol{x}}_{\rm tr}(t_k)=\hat{\boldsymbol{x}}(t_k)$, $k=2,3,\cdots,n$.

Proof. By substituting $\boldsymbol{y}_{\rm tr}(t_k)=\boldsymbol{y}(t_k)+\boldsymbol{\xi}$ into the matrices in equation (9), the two matrices therein are rewritten as

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

and

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

From equation (11) one has the estimated parameters expressed as

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

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

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

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

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

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

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

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

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

and the restored fitting and forecasting values

$$\hat{\boldsymbol{x}}_{\rm tr}(t_k) = \frac{\hat{\boldsymbol{y}}_{\rm tr}(t_k) - \hat{\boldsymbol{y}}_{\rm tr}(t_{k-1})}{t_k - t_{k-1}} = \frac{\hat{\boldsymbol{y}}(t_k) - \hat{\boldsymbol{y}}(t_{k-1})}{t_k - t_{k-1}} = \boldsymbol{x}(t_k),\ k = 2,3,\cdots,n+r.$$

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 ($\boldsymbol{X}_{\rm tr}(t) = \{\boldsymbol{x}(t_1)+\boldsymbol{\xi}, \boldsymbol{x}(t_2),\cdots,\boldsymbol{x}(t_n)\}$). Thus, without changing performance, the Cusum operator in Definition 1 can be defined also as $\boldsymbol{y}(t_1) = \boldsymbol{\xi}$ and $\boldsymbol{y}(t_k) = \boldsymbol{\xi} + \sum_{i=2}^{k} h_k\boldsymbol{x}(t_k)$ when $k \geq 2$. □

Theorem 1. *Let*

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

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

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

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

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

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

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

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

Conversely, integrating both sides of equation (18) gives

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

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

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

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

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

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

Theorem 1 shows that by introducing the integral transformation formula (17), equation (6) can always be reduced to the simple form (18) and, in the other direction, equation (18) can express (6) by selecting an appropriate initial value $\boldsymbol{x}(t_1) = \boldsymbol{c} + \boldsymbol{B}\boldsymbol{u}(t_1) + \boldsymbol{A}\boldsymbol{\xi}$. In this sense, equation (6) has a redundant degree of freedom (the parameter vector $\boldsymbol{c}$) and it can be eliminated by using the reduced-order equation (18) to model the original time series directly.

Example 1. *Supposing the ordinary differential equation* (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*

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

Then the equivalent reduced-order equation (18) *is* $\frac{d}{dt}x(t) = ax(t) + 2b_2t + b_1,\ x(0) = c + a\xi,\ t \geq 0$, *whose closed-form solution is*

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

and it is easy to validate that $y(t) = \xi + \int_0^t x(s)ds$, *which is consistent with equation* (17).

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

3.2. Integral transformation for explaining and improving the cumulative sum operator

The integral transformation formula (17) plays an important role in the simplification from equation (6) to (18). Since the observations of the state vector $\boldsymbol{x}(t)$ at time points $\{t_k\}_{k=1}^{n}$ are known in advance, the

finite integral in equation (17) can be calculated by using the piecewise interpolation methods, such as the piecewise constant integration formula

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

and the piecewise linear integration formula

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

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

Note that equation (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 equation (17), such as the Simpson and Gauss-Legendre rules.

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

Integral matching-based model uses the reduced-order equation (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 equation (18) gives the definite integral equation (20). By substituting $\boldsymbol{x}(t_1) = \boldsymbol{\eta}$ into equation (20), one has

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

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

$$\boldsymbol{x}(t_k) = \boldsymbol{A}\left[\boldsymbol{y}_{\text{lnt}}(t_k) - \boldsymbol{x}(t_1)\right] + \boldsymbol{B}\left[\boldsymbol{u}(t_k) - \boldsymbol{u}(t_1)\right] + \boldsymbol{\eta} + \boldsymbol{\epsilon}(t_k) \tag{23}$$

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

Similar to the treatment process in section 2, by substituting $k = 2, 3, \cdots, n$ and the observed data into equation (23), and then arranging the obtained $n-1$ equations into a matrix form, one has

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

where

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

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

$$\hat{\boldsymbol{\Pi}} = \left(\boldsymbol{\Omega}^{\mathsf{T}}(\boldsymbol{x}, \boldsymbol{u})\boldsymbol{\Omega}(\boldsymbol{x}, \boldsymbol{u})\right)^{-1}\boldsymbol{\Omega}^{\mathsf{T}}(\boldsymbol{x}, \boldsymbol{u})\boldsymbol{X}. \tag{25}$$

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

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

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 equation (26).

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

In this section, we analyze the relationship between grey forecasting models and integral matching-based differential equation models from both the parameter estimation and modelling procedure perspectives.

For convenience, the original time series is assumed to be equally spaced, i.e., $h_k = h$, $k = 2, 3, \cdots, n$. The estimated parameters of grey models are denoted as $\hat{\boldsymbol{A}}_{\mathrm{g}}$, $\hat{\boldsymbol{B}}_{\mathrm{g}}$ and $\hat{\boldsymbol{c}}_{\mathrm{g}}$, and those of integral matching-based ones are denoted as $\hat{\boldsymbol{A}}_{\mathrm{m}}$, $\hat{\boldsymbol{B}}_{\mathrm{m}}$ and $\hat{\boldsymbol{\eta}}_{\mathrm{m}}$.

4.1. Quantitative relationship between the estimated parameters

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

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

$$\begin{bmatrix}\hat{\boldsymbol{A}}_{\mathrm{g}}^{\mathsf{T}}\\ \hat{\boldsymbol{c}}_{\mathrm{g}}^{\mathsf{T}}\end{bmatrix} = \left(\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}}\right)^{-1}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{X} \quad \text{and} \quad \begin{bmatrix}\hat{\boldsymbol{A}}_{\mathrm{m}}^{\mathsf{T}}\\ \hat{\boldsymbol{\eta}}_{\mathrm{m}}^{\mathsf{T}}\end{bmatrix} = \left(\boldsymbol{Y}_{\mathrm{m}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{m}}\right)^{-1}\boldsymbol{Y}_{\mathrm{m}}^{\mathsf{T}}\boldsymbol{X}$$

where

$$\boldsymbol{Y}_{\mathrm{g}} = \begin{bmatrix}\frac{\boldsymbol{y}^{\mathsf{T}}(t_1)+\boldsymbol{y}^{\mathsf{T}}(t_2)}{2} & 1\\ \frac{\boldsymbol{y}^{\mathsf{T}}(t_2)+\boldsymbol{y}^{\mathsf{T}}(t_3)}{2} & 1\\ \vdots & \vdots\\ \frac{\boldsymbol{y}^{\mathsf{T}}(t_{n-1})+\boldsymbol{y}^{\mathsf{T}}(t_n)}{2} & 1\end{bmatrix} \quad \text{and} \quad \boldsymbol{Y}_{\mathrm{m}} = \begin{bmatrix}\boldsymbol{y}_{\mathrm{lnt}}^{\mathsf{T}}(t_2) - \boldsymbol{x}^{\mathsf{T}}(t_1) & 1\\ \boldsymbol{y}_{\mathrm{lnt}}^{\mathsf{T}}(t_3) - \boldsymbol{x}^{\mathsf{T}}(t_1) & 1\\ \vdots & \vdots\\ \boldsymbol{y}_{\mathrm{lnt}}^{\mathsf{T}}(t_n) - \boldsymbol{x}^{\mathsf{T}}(t_1) & 1\end{bmatrix}.$$

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

$$\boldsymbol{y}_{\mathrm{lnt}}(t_k) = \frac{1}{2}\boldsymbol{y}(t_{k-1}) + \frac{1}{2}\boldsymbol{y}(t_k) + \frac{h}{2}\boldsymbol{x}(t_1),\ k = 2, 3, \cdots, n$$

and thus we have $\boldsymbol{Y}_{\mathrm{m}} = \boldsymbol{Y}_{\mathrm{g}}\boldsymbol{Q}$, where $\boldsymbol{I}_d$ is the dth-order unit matrix,

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

Finally, a simple substitution gives the desired equation

$$\begin{bmatrix}\hat{\boldsymbol{A}}_{\mathrm{m}}^{\mathsf{T}}\\ \hat{\boldsymbol{\eta}}_{\mathrm{m}}^{\mathsf{T}}\end{bmatrix} = \left(\left(\boldsymbol{Y}_{\mathrm{g}}\boldsymbol{Q}\right)^{\mathsf{T}}\left(\boldsymbol{Y}_{\mathrm{g}}\boldsymbol{Q}\right)\right)^{-1}\left(\boldsymbol{Y}_{\mathrm{g}}\boldsymbol{Q}\right)^{\mathsf{T}}\boldsymbol{X} = \boldsymbol{Q}^{-1}\begin{bmatrix}\hat{\boldsymbol{A}}_{\mathrm{g}}^{\mathsf{T}}\\ \hat{\boldsymbol{c}}_{\mathrm{g}}^{\mathsf{T}}\end{bmatrix} = \begin{bmatrix}\hat{\boldsymbol{A}}_{\mathrm{g}}^{\mathsf{T}}\\ \hat{\boldsymbol{c}}_{\mathrm{g}}^{\mathsf{T}} + \boldsymbol{x}^{\mathsf{T}}(t_1)\hat{\boldsymbol{A}}_{\mathrm{g}}^{\mathsf{T}} - \frac{h}{2}\boldsymbol{x}^{\mathsf{T}}(t_1)\hat{\boldsymbol{A}}_{\mathrm{g}}^{\mathsf{T}}\end{bmatrix}.$$

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

If the independent vector $\boldsymbol{u}(t) \neq 0$, the quantitative relationship is not that obvious. By partitioning the estimated parameter matrices in equations (11) and (25) as $\hat{\boldsymbol{\Xi}}_{\mathrm{g}} = \left[\hat{\boldsymbol{A}}_{\mathrm{g}}\ \hat{\boldsymbol{c}}_{\mathrm{g}} \mid \hat{\boldsymbol{B}}_{\mathrm{g}}\right]$ and $\hat{\boldsymbol{\Pi}}_{\mathrm{m}} = \left[\hat{\boldsymbol{A}}_{\mathrm{m}}\ \hat{\boldsymbol{\eta}}_{\mathrm{m}} \mid \hat{\boldsymbol{B}}_{\mathrm{m}}\right]$, then we have

$$\hat{\boldsymbol{\Xi}}_{\mathrm{g}} = \left(\begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{U}_{\mathrm{g}}\end{bmatrix}\right)^{-1}\begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\end{bmatrix}\boldsymbol{X} = \begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{X} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{X}\end{bmatrix} \tag{27}$$

and

$$\hat{\boldsymbol{\Pi}}_{\mathrm{m}} = \left(\begin{bmatrix}\boldsymbol{Y}_{\mathrm{m}}^{\mathsf{T}} \\ \boldsymbol{U}_{\mathrm{m}}^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}\boldsymbol{Y}_{\mathrm{m}} & \boldsymbol{U}_{\mathrm{m}}\end{bmatrix}\right)^{-1}\begin{bmatrix}\boldsymbol{Y}_{\mathrm{m}}^{\mathsf{T}} \\ \boldsymbol{U}_{\mathrm{m}}^{\mathsf{T}}\end{bmatrix}\boldsymbol{X} = \begin{bmatrix}\boldsymbol{Y}_{\mathrm{m}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{m}} & \boldsymbol{Y}_{\mathrm{m}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{m}} \\ \boldsymbol{U}_{\mathrm{m}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{m}} & \boldsymbol{U}_{\mathrm{m}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{m}}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{Y}_{\mathrm{m}}^{\mathsf{T}}\boldsymbol{X} \\ \boldsymbol{U}_{\mathrm{m}}^{\mathsf{T}}\boldsymbol{X}\end{bmatrix} \tag{28}$$

where

$$\boldsymbol{U}_{\mathrm{g}} = \frac{1}{2}\begin{bmatrix}\boldsymbol{u}^{\mathsf{T}}(t_2)+\boldsymbol{u}^{\mathsf{T}}(t_1) \\ \boldsymbol{u}^{\mathsf{T}}(t_3)+\boldsymbol{u}^{\mathsf{T}}(t_2) \\ \vdots \\ \boldsymbol{u}^{\mathsf{T}}(t_n)+\boldsymbol{u}^{\mathsf{T}}(t_{n-1})\end{bmatrix} = \begin{bmatrix}\boldsymbol{u}^{\mathsf{T}}(t_2) \\ \boldsymbol{u}^{\mathsf{T}}(t_3) \\ \vdots \\ \boldsymbol{u}^{\mathsf{T}}(t_n)\end{bmatrix} + \frac{1}{2}\begin{bmatrix}\boldsymbol{u}^{\mathsf{T}}(t_1)-\boldsymbol{u}^{\mathsf{T}}(t_2) \\ \boldsymbol{u}^{\mathsf{T}}(t_2)-\boldsymbol{u}^{\mathsf{T}}(t_3) \\ \vdots \\ \boldsymbol{u}^{\mathsf{T}}(t_{n-1})-\boldsymbol{u}^{\mathsf{T}}(t_n)\end{bmatrix}$$

and

$$\boldsymbol{U}_{\mathrm{m}} = \begin{bmatrix}\boldsymbol{u}^{\mathsf{T}}(t_2)-\boldsymbol{u}^{\mathsf{T}}(t_1) \\ \boldsymbol{u}^{\mathsf{T}}(t_3)-\boldsymbol{u}^{\mathsf{T}}(t_1) \\ \vdots \\ \boldsymbol{u}^{\mathsf{T}}(t_n)-\boldsymbol{u}^{\mathsf{T}}(t_1)\end{bmatrix} = \begin{bmatrix}\boldsymbol{u}^{\mathsf{T}}(t_2) \\ \boldsymbol{u}^{\mathsf{T}}(t_3) \\ \vdots \\ \boldsymbol{u}^{\mathsf{T}}(t_n)\end{bmatrix} - \begin{bmatrix}\boldsymbol{u}^{\mathsf{T}}(t_1) \\ \boldsymbol{u}^{\mathsf{T}}(t_1) \\ \vdots \\ \boldsymbol{u}^{\mathsf{T}}(t_1)\end{bmatrix} = \boldsymbol{U}_{\mathrm{g}} + \frac{1}{2}\begin{bmatrix}\boldsymbol{u}^{\mathsf{T}}(t_2)-\boldsymbol{u}^{\mathsf{T}}(t_1) \\ \boldsymbol{u}^{\mathsf{T}}(t_3)-\boldsymbol{u}^{\mathsf{T}}(t_2) \\ \vdots \\ \boldsymbol{u}^{\mathsf{T}}(t_n)-\boldsymbol{u}^{\mathsf{T}}(t_{n-1})\end{bmatrix} - \begin{bmatrix}\boldsymbol{u}^{\mathsf{T}}(t_1) \\ \boldsymbol{u}^{\mathsf{T}}(t_1) \\ \vdots \\ \boldsymbol{u}^{\mathsf{T}}(t_1)\end{bmatrix}.$$

By denoting $\boldsymbol{U}_{\mathrm{m}} = \boldsymbol{U}_{\mathrm{g}} + \boldsymbol{U}$ and combining $\boldsymbol{Y}_{\mathrm{m}} = \boldsymbol{Y}_{\mathrm{g}}\boldsymbol{Q}$ in Proposition 1, then the matrices in equation (28) can be rewritten as

$$\begin{bmatrix}\boldsymbol{Y}_{\mathrm{m}}^{\mathsf{T}}\boldsymbol{X} \\ \boldsymbol{U}_{\mathrm{m}}^{\mathsf{T}}\boldsymbol{X}\end{bmatrix} = \begin{bmatrix}\boldsymbol{Q}^{\mathsf{T}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_p\end{bmatrix}\left(\begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{X} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{X}\end{bmatrix} + \begin{bmatrix}\boldsymbol{0} \\ \boldsymbol{U}^{\mathsf{T}}\boldsymbol{X}\end{bmatrix}\right) \tag{29}$$

and

$$\begin{bmatrix}\boldsymbol{Y}_{\mathrm{m}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{m}} & \boldsymbol{Y}_{\mathrm{m}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{m}} \\ \boldsymbol{U}_{\mathrm{m}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{m}} & \boldsymbol{U}_{\mathrm{m}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{m}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{Q}^{\mathsf{T}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_p\end{bmatrix}\left(\begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}}\end{bmatrix} + \begin{bmatrix}\boldsymbol{0} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U} \\ \boldsymbol{U}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{W}\end{bmatrix}\right)\begin{bmatrix}\boldsymbol{Q} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_p\end{bmatrix} \tag{30}$$

where $\boldsymbol{W} = \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U} + \boldsymbol{U}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}} + \boldsymbol{U}^{\mathsf{T}}\boldsymbol{U}$.

By substituting equations (29) and (30) into equation (28), the estimates are obtained as

$$\hat{\boldsymbol{\Pi}}_{\mathrm{m}} = \begin{bmatrix}\boldsymbol{Q}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_p\end{bmatrix}\left(\begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}}\end{bmatrix} + \begin{bmatrix}\boldsymbol{0} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U} \\ \boldsymbol{U}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{W}\end{bmatrix}\right)^{-1}\left(\begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{X} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{X}\end{bmatrix} + \begin{bmatrix}\boldsymbol{0} \\ \boldsymbol{U}^{\mathsf{T}}\boldsymbol{X}\end{bmatrix}\right) \tag{31}$$

where, by using the Sherman-Morrison-Woodbury formula [69], the inverse matrix can be rewritten as

$$\begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}}\end{bmatrix}^{-1} - \begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}}\end{bmatrix}^{-1}\left(\begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}}\end{bmatrix}^{-1} + \begin{bmatrix}\boldsymbol{0} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U} \\ \boldsymbol{U}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{W}\end{bmatrix}^{-1}\right)\begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}}\end{bmatrix}^{-1}.$$

Therefore, it follows that

$$\begin{aligned}\hat{\boldsymbol{\Pi}}_{\mathrm{m}} = {} & \begin{bmatrix}\boldsymbol{Q}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_p\end{bmatrix}\hat{\boldsymbol{\Xi}}_{\mathrm{g}} \\ & + \begin{bmatrix}\boldsymbol{Q}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_p\end{bmatrix}\begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{0} \\ \boldsymbol{U}^{\mathsf{T}}\boldsymbol{X}\end{bmatrix} \\ & - \begin{bmatrix}\boldsymbol{Q}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_p\end{bmatrix}\begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}}\end{bmatrix}^{-1}\left(\begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}}\end{bmatrix}^{-1} + \begin{bmatrix}\boldsymbol{0} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U} \\ \boldsymbol{U}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{W}\end{bmatrix}^{-1}\right)\hat{\boldsymbol{\Xi}}_{\mathrm{g}} \\ & - \begin{bmatrix}\boldsymbol{Q}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_p\end{bmatrix}\begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}}\end{bmatrix}^{-1}\left(\begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}}\end{bmatrix}^{-1} + \begin{bmatrix}\boldsymbol{0} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U} \\ \boldsymbol{U}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{W}\end{bmatrix}^{-1}\right)\begin{bmatrix}\boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{Y}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}} \\ \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{Y}_{\mathrm{g}} & \boldsymbol{U}_{\mathrm{g}}^{\mathsf{T}}\boldsymbol{U}_{\mathrm{g}}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{0} \\ \boldsymbol{U}^{\mathsf{T}}\boldsymbol{X}\end{bmatrix}\end{aligned}$$

where the first term on the right hand side shows that

$$\hat{\boldsymbol{A}}_{\mathrm{g}} = \hat{\boldsymbol{A}}_{\mathrm{m}},\ \hat{\boldsymbol{B}}_{\mathrm{g}} = \hat{\boldsymbol{B}}_{\mathrm{m}},\ \hat{\boldsymbol{\eta}}_{\mathrm{m}} = \hat{\boldsymbol{c}}_{\mathrm{g}} - \frac{h-2}{2}\hat{\boldsymbol{A}}_{\mathrm{g}}\boldsymbol{x}(t_1)$$

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

4.2. Modelling procedure 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.

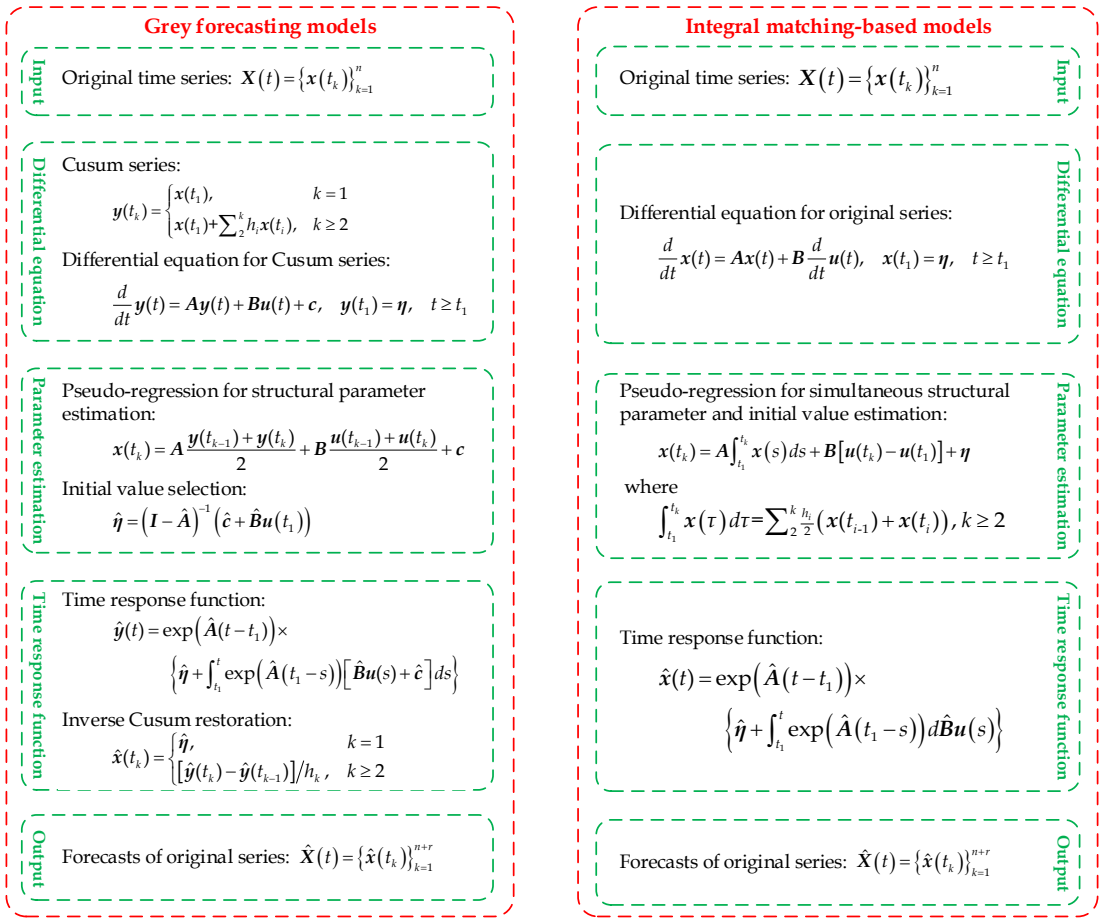

Figure 2: Reconstructed modelling procedure of grey models by integral matching-based models.

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

(i) Grey models use complex 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 in order to estimate 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 equation (15), which may well introduce additional errors [70].

(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 $u_\iota(t)$ is known in advance, such as the deterministic function of time in Remark 1. In this case, there exists time-lag effect between the state variable and independent variable, such as the cases in Remark 2 having $u_\iota(t) = y_\iota(t - \tau_\iota h)$ where $\tau_\iota \geq 1$ 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 forecast over the forecasting horizon. For example, univariate or multivariate time series methods can be applied, such as the dynamic harmonic regression [71, 72]; see [3] for a comprehensive example.

5. Simulation studies

In this section, we compare grey models with integral matching-based models by large-scale Monte Carlo simulation studies associated with the statistical comparison of the simulation results.

The time-series data are generated from the following observation equation

$$\tilde{\boldsymbol{x}}(t_k) = \begin{cases} \boldsymbol{x}(t_k) + \boldsymbol{e}(k), & k = 1, 2, \cdots, n \quad \leftarrow \text{in-sample time series for training} \\ \boldsymbol{x}(t_k), & k = n+1, \cdots, n+r \quad \leftarrow \text{out-of-sample time series for validating} \end{cases} \tag{32}$$

where $\boldsymbol{x}(t)$ is the solution to the state equation (18); $\boldsymbol{e}(k) \sim \mathcal{N}(0, \boldsymbol{\Sigma})$ is a d vector normal distribution with mean 0 and covariance matrix $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_d$.

The fitting and forecasting errors are measured by the in-sample and out-ot-sample mean absolute percentage error criteria, respectively expressed as

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

where $x_\ell(t)$, $\ell = 1, 2, \cdots, d$, is the ℓth component of the state variable $\boldsymbol{x}(t)$.

In order to analyze the effect of sample size and noise level on modelling, the sample size n and standard deviation σ are set as different values, i.e, the data are sampled at every time interval of h in the range of $t \in [0, T]$, thereby generating $n = \left[\frac{T}{h}\right] + 1$ samples; and the standard deviation is $\sigma = \frac{1}{\sqrt{snr}}\sqrt{\text{Var}(x_\ell(t))}$, where snr is the signal-to-noise ratio used to measure the noise level. 1000 Monte Carlo realizations are repeated for each sample size and signal-to-noise ratio combination. Figure 3 shows the calculation of the performance evaluation indexes in each run.

5.1. Case 1: the single-output model

Consider the case where the dimension of state variable is $d = 1$ and the independent vector consists of $u_1(t) = t$ and $u_2(t) = t^2$. Let the true parameters be $a = 0.15$, $b_2 = 0.10$, $b_1 = -0.25$ and $\eta = 1.20$. Then equation (18) is

$$\frac{d}{dt}x(t) = ax(t) + 2b_2 t + b_1 = 0.15x(t) + 0.20t - 0.25, \ t \geq 0$$

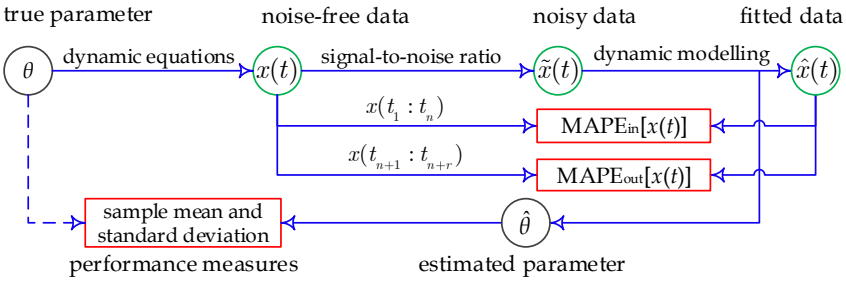

Figure 3: Diagrammatic representation of simulation design and performance evaluation in each run.

where the initial value is $x(0) = \eta = 1.20$; corresponding, by Theorem 1 equation (6) is expressed as

$$\frac{d}{dt}y(t) = ay(t) + b_2t^2 + b_1t + c = 0.15y(t) + 0.10t^2 - 0.25t + 1.02,\ t \geq 0$$

where the initial value is $y(0) = \eta = 1.20$.

Taking the time interval as $h \in [0.25, 0.10, 0.05]$ and the time upper bound as $T = 5.0$, then the sample size is $n \in [21, 51, 101]$. Let the signal-to-noise ratio be $snr \in [2.5, 3.5, 5.0]$. Following the data generation in equation (32), the simulation data for $3 \times 3 = 9$ scenarios are obtained and the modelling results of integral matching and grey modelling approaches are summarized in Table 2 and Figure 4.

Table 2: Sample means (sample standard deviations) of the estimated parameters obtained from grey modelling and integral matching 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 $a = 0.15$, $b_2 = 0.10$, $b_1 = -0.25$ and $\eta = 1.20$.

n	snr	Grey modelling [a]				Integral matching [b]			
		$\hat{a}$	$\hat{b}_2$	$\hat{b}_1$	$\hat{\eta}$	$\hat{a}$	$\hat{b}_2$	$\hat{b}_1$	$\hat{\eta}$
21	2.5	0.116(0.631)	0.122(0.185)	−0.270(0.433)	1.404(4.611)	0.116(0.631)	0.122(0.185)	−0.300(0.407)	1.272(0.348)
	3.5	0.145(0.378)	0.105(0.108)	−0.244(0.241)	1.195(1.135)	0.145(0.378)	0.105(0.108)	−0.270(0.222)	1.223(0.189)
	5.0	0.150(0.187)	0.101(0.054)	−0.231(0.117)	1.181(0.114)	0.150(0.187)	0.101(0.054)	−0.256(0.108)	1.206(0.094)
51	2.5	0.152(0.268)	0.102(0.073)	−0.253(0.176)	1.227(1.492)	0.152(0.268)	0.102(0.073)	−0.264(0.170)	1.210(0.120)
	3.5	0.152(0.138)	0.100(0.038)	−0.245(0.089)	1.191(0.081)	0.152(0.138)	0.100(0.038)	−0.255(0.086)	1.203(0.062)
	5.0	0.151(0.068)	0.100(0.019)	−0.242(0.043)	1.195(0.036)	0.151(0.068)	0.100(0.019)	−0.252(0.042)	1.201(0.030)
101	2.5	0.141(0.150)	0.102(0.041)	−0.240(0.095)	1.194(0.100)	0.141(0.150)	0.102(0.041)	−0.245(0.093)	1.203(0.064)
	3.5	0.146(0.076)	0.101(0.021)	−0.242(0.048)	1.198(0.042)	0.146(0.076)	0.101(0.021)	−0.247(0.047)	1.201(0.033)
	5.0	0.148(0.037)	0.100(0.010)	−0.243(0.023)	1.198(0.020)	0.148(0.037)	0.100(0.010)	−0.248(0.023)	1.200(0.016)

[a] In grey modelling the estimated initial value is computed on the basis of $\hat{\eta} = \frac{\hat{c}}{1-\hat{a}}$. [b] In integral matching the estimated parameter $\hat{b}_2$ is divided by 2 in order to make the estimates comparable, i.e., $\hat{b}_2 = \frac{1}{2}\hat{b}_{2_m}$.

Table 2 shows that with the increase of the signal-to-noise ratio, the sample means of the estimates go to the true values and in the meantime, their sample standard deviations get smaller. In detail, grey modelling and integral matching have the same estimates $\hat{a}$ and $\hat{b}_2$; the estimate $\hat{b}_1$ obtained from grey modelling is a little closer to the true value $b_1 = -0.25$ than that obtained from integral matching in the small-sample size cases ($n = 21$), whereas the former has much larger sample standard deviations than the latter, especially in the small-sample size cases ($n = 21$); the estimated initial value $\hat{\eta}$ obtained from integral matching is much closer to the true initial value $\eta = 1.20$ and also has much smaller sample standard deviations than that obtained from grey modelling. In summary, integral matching has better parameter estimation performance in terms of the accuracy (measured by the sample mean) and robustness to noise (measured by the sample standard deviation). Besides, 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 (the details will be discussed in section 7.2).

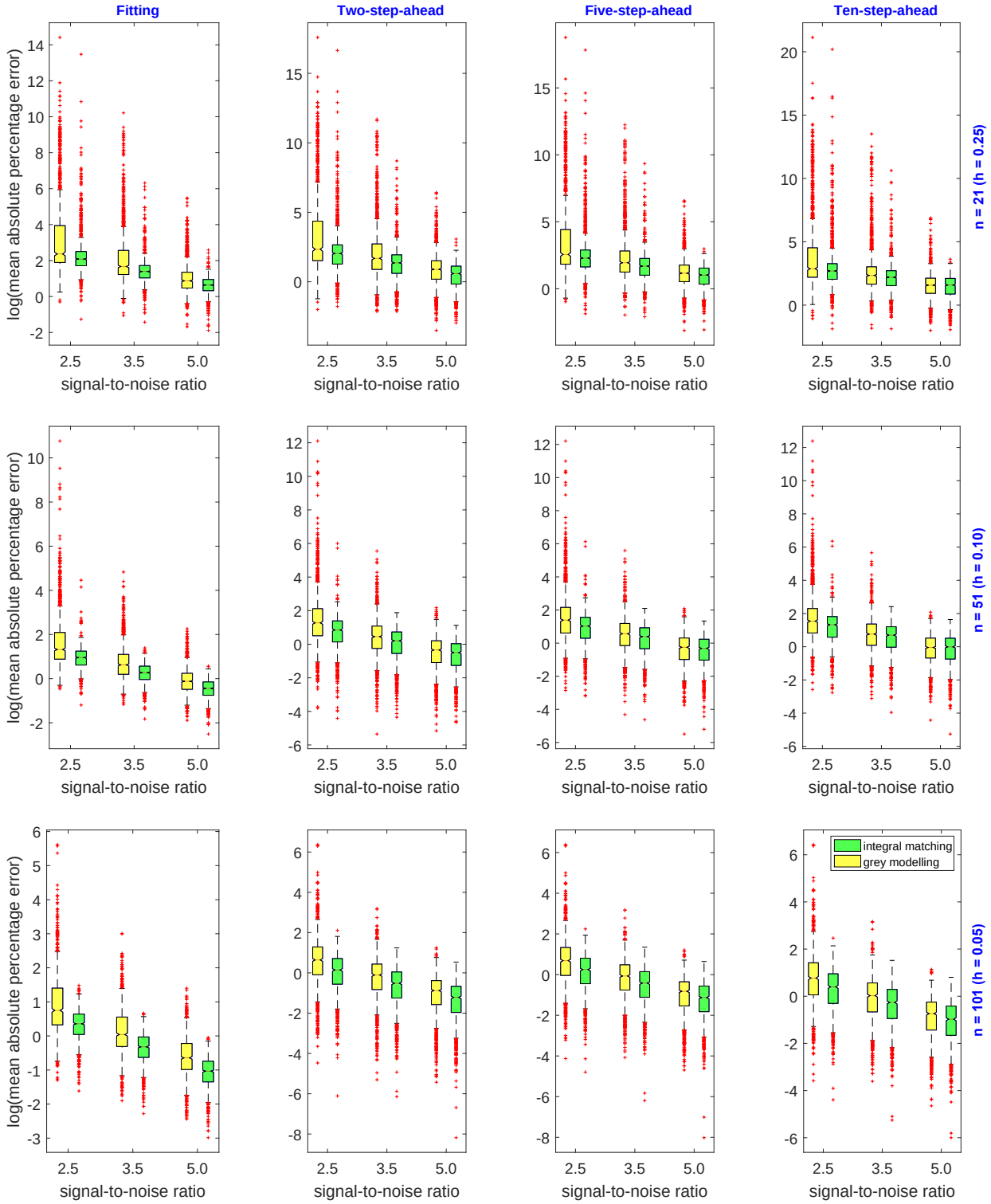

Figure 4: Boxplots of the fitting and multi-step-ahead forecasting errors obtained form two approaches in 1000 replications. Note that the y-axis is reset to be logarithmic coordinate to make the boxplots comparable (due to that the maximums of the outliers obtained from grey modelling are too large in the original coordinate).

Figure 4 shows that integral matching has much smaller fitting and multi-step-ahead forecasting errors than grey modelling for all the sample size and signal-to-noise ratio combination scenarios, indicating that the former outperforms the latter from both fitting and forecasting viewpoints. The fitting and forecasting errors vary a lot when the signal-to-noise ratio is low ($snr = 2.5, 3.5$) in the small-sample size scenarios ($n = 21$), indicating that we need to be careful when using these two approaches to solve the small-sample forecasting problems. Fortunately, we can increase the sample size or the signal-to-noise ratio to improve accuracy and robustness.

Since Figure 4 only depicts the brief distributions of the fitting and forecasting errors, the paired T-test is also employed to compare the fitting and forecasting errors from a statistical significance viewpoint. The alternative hypothesis is

$$\mu_{\mathrm{g}}(\mathcal{I}) - \mu_{\mathrm{m}}(\mathcal{I}) > 0$$

where the index $\mathcal{I}$ is $\mathrm{MAPE}_{\mathrm{in}}$ or $\mathrm{MAPE}_{\mathrm{out}}$; $\mu_{\mathrm{g}}(\mathcal{I})$ and $\mu_{\mathrm{m}}(\mathcal{I})$ are means of $\mathcal{I}$ obtained from grey modelling and integral matching, respectively. Table 3 shows that as the sample size (or signal-to-noise ratio) increases, (i) the mean of the differences between fitting (or multi-step-ahead forecasting) errors decrease sharply, indicating that grey modelling and integral matching tend to obtain similar performance in the high signal-to-noise ratio or large sample size cases; (ii) although the mean of the differences gets smaller, the means of $\mathrm{MAPE}_{\mathrm{in}}$s and $\mathrm{MAPE}_{\mathrm{out}}$s obtained from grey modelling are significantly greater than those obtained from integral matching, respectively, indicating that integral matching gains significant advantages over grey modelling in terms of both fitting and multi-step-ahead forecasting.

Table 3: Statistical comparison (mean of the differences and significance) of the fitting and multi-step-ahead forecasting errors of grey modelling and integral matching approaches in 1000 runs with the sample size $n = 21, 51, 101$ and signal-to-noise ratio $snr = 2.5, 3.5, 5.0$.

sample	Fitting			Two-step-ahead forecasting			Five-step-ahead forecasting			Ten-step-ahead forecasting		
size	2.5	3.5	5.0	2.5	3.5	5.0	2.5	3.5	5.0	2.5	3.5	5.0
21	2.0e3	1.6e2	3.39	3.3e4	7.8e2	5.82	9.9e4	1.1e3	5.89	9.7e5	2.4e3	6.07
	0.034*	0.000**	0.000**	0.106	0.000**	0.000**	0.127	0.000**	0.000**	0.146	0.002*	0.000**
51	1.0e2	2.06	0.45	3.7e2	2.34	0.22	4.1e2	2.21	0.15	5.0e2	1.98	0.06
	0.023*	0.000**	0.000**	0.028*	0.000**	0.000**	0.028*	0.000**	0.000**	0.028*	0.000**	0.002**
101	3.42	0.74	0.28	4.58	0.52	0.14	4.57	0.48	0.12	4.54	0.42	0.10
	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**

Note that the asterisks * and ** indicate significance at the level 5% and 1%.

5.2. Case 2: the multi-output model

Consider the case with the dimensions $d = 2$ and $p = 0$. Let the true structural parameter matrix and initial vector be

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

Then we have equation (18) expressed as

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

and equation (20), from Theorem 1, expressed as

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

The parameters are set to the same values as those in section 5.1, i.e., $T = 5.0$, $h \in [0.25, 0.10, 0.05]$ ($n \in [21, 51, 101]$), and $snr \in [2.5, 3.5, 5.0]$ and, similar to the modelling procedures in section 5.1, the results obtained from integral matching and grey modelling are summarized in Table 4 and Figures 5 and 6.

Table 4 shows that (i) the estimated structural parameter matrices obtained from grey modelling and integral matching are equal to each other, validating the Proposition 1 in section 4.1; (ii) the estimates of the parameters, including structural parameters and initial values, tend to their corresponding true values, and meanwhile, the sample standard deviations of the estimates get smaller with the increase of signal-to-noise ratio, partially implying the asymptotic convergence of these two parameter estimation approaches; (iii) 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 than grey modelling.

Figures 5 and 6 show that the differences between the fitting/forecasting errors are extremely significant, so it is not necessary to employ the paired T-test to compare the results. The error distributions of $x_1(t)$ in Figure 5 and $x_2(t)$ in Figure 6 show that the fitting and multi-step-ahead forecasting errors obtained from grey modelling are much greater than those obtained from integral matching in all the sample size and signal-to-noise ratio combination scenarios, indicating that integral matching outperforms grey modelling from both fitting and multi-step-ahead forecasting viewpoints.

Table 4: 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				Initial vector [a]		Initial vector [b]	
		$\hat{a}_{1,1}$	$\hat{a}_{1,2}$	$\hat{a}_{2,1}$	$\hat{a}_{2,2}$	$\hat{\eta}_1$	$\hat{\eta}_2$	$\hat{\eta}_1$	$\hat{\eta}_2$
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)

[a] The estimated initial vector of grey modelling is obtained from $\hat{\boldsymbol{\eta}} = \left(\boldsymbol{I}_2 - \hat{\boldsymbol{A}}\right)^{-1} \hat{\boldsymbol{c}}$. [b] The estimated initial vector corresponds to integral matching.

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 the two main water supply sources. Facing the global groundwater crisis [73], 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 convenience, 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 ($10^9 \mathsf{m}^3$) 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 5.

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)

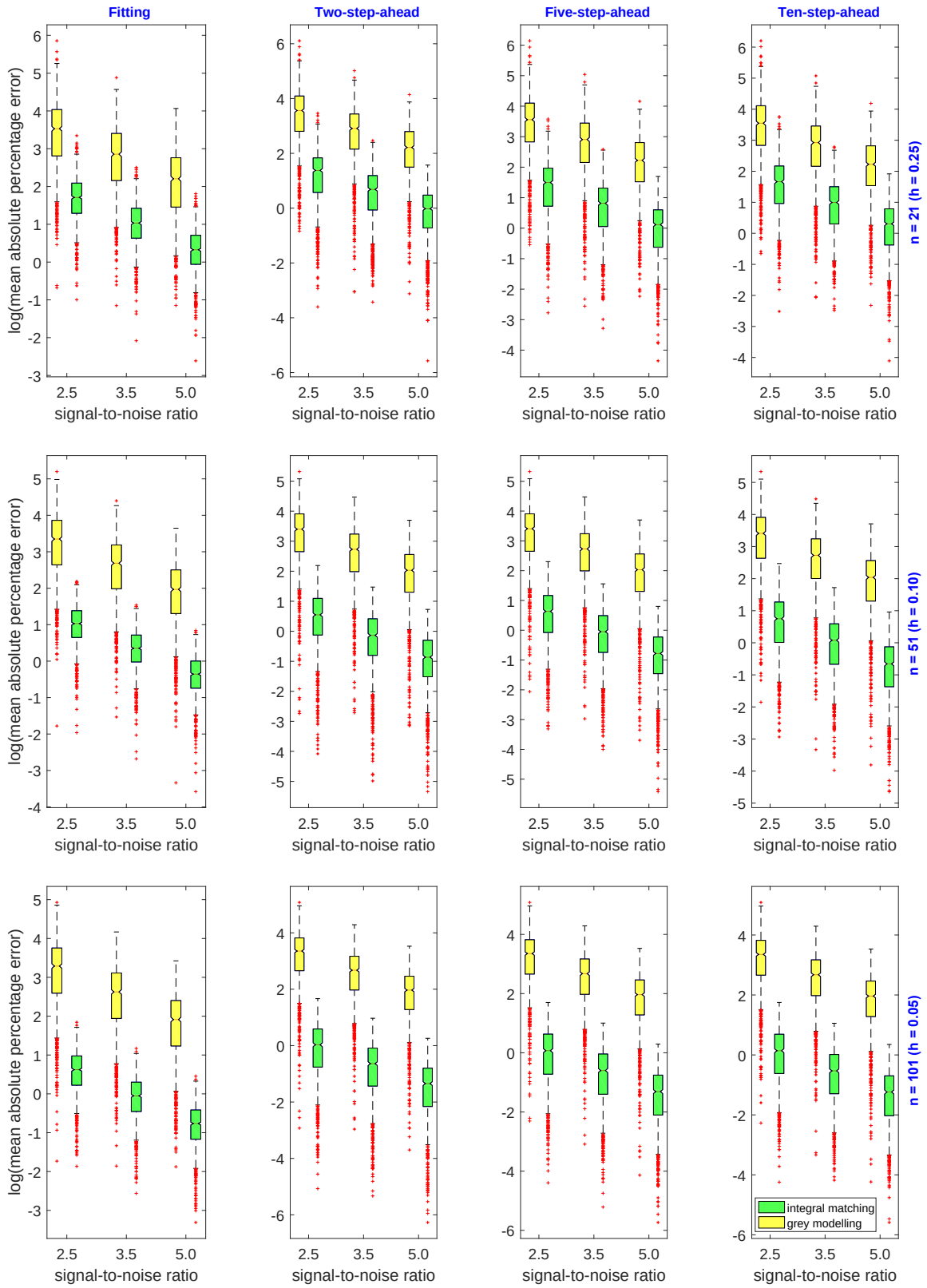

Figure 5: Boxplots of the fitting and multi-step-ahead forecasting errors of the first component $x_1(t)$ in 1000 replications.

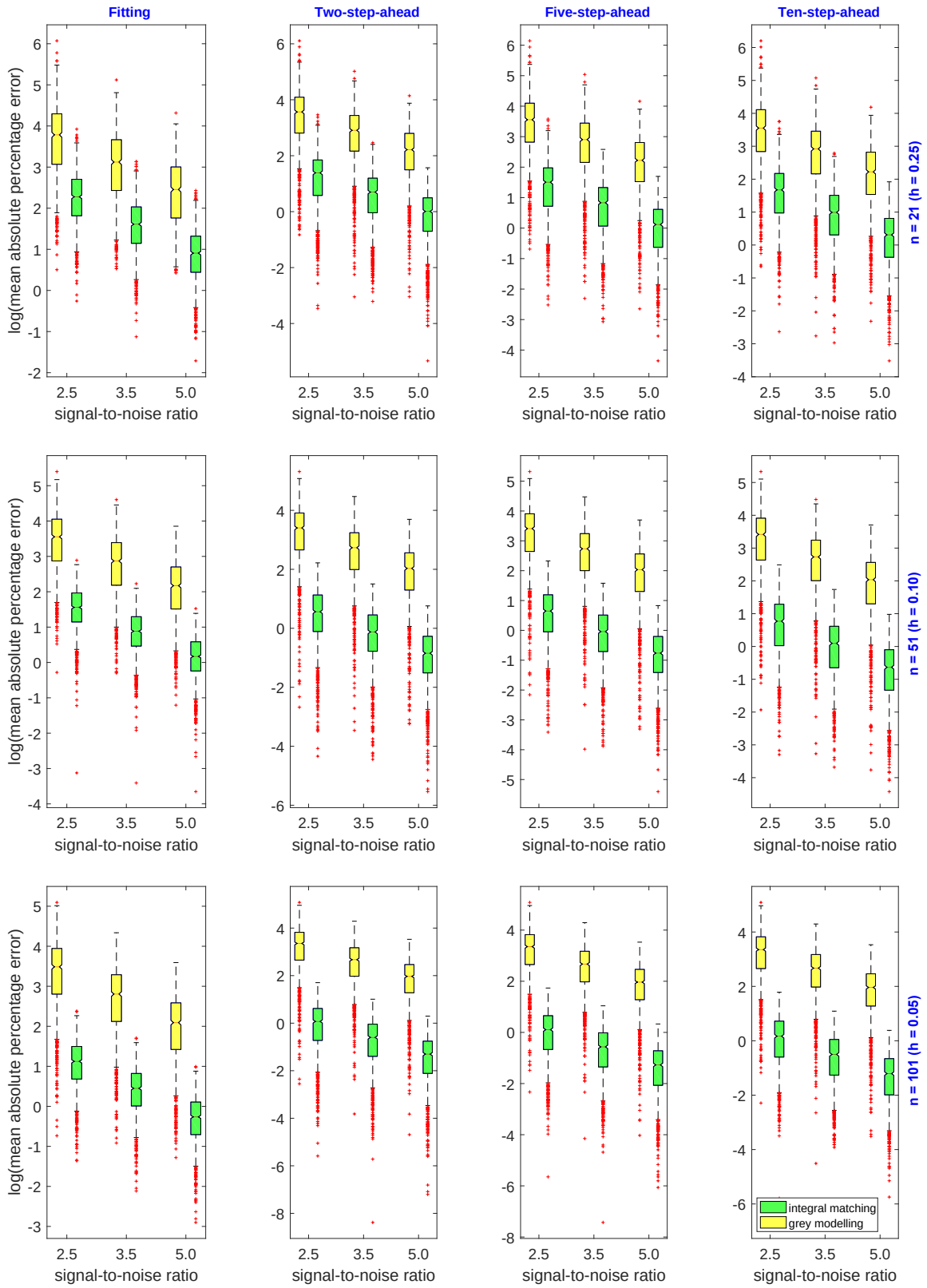

Figure 6: Boxplots of the fitting and multi-step-ahead forecasting errors of the second component $x_2(t)$ in 1000 replications.

Table 5: Fitting and forecasting results obtained from integral matching-based and grey models with variant model structures.

Year	True	Model 1 [i]		Model 2 [ii]		Model 3 [iii]		Model 4 [iv]		Model 5 [v]		Model 6 [vi]	
	value	Value	APE	Value	APE	Value	APE	Value	APE	Value	APE	Value	APE
2004	17.20	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	21.96	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	22.70	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	25.70	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	28.74	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	31.16	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	33.12	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	44.80	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	44.60	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	49.94	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	57.46	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	64.50	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_{in} (%)			3.46		3.90		4.40		4.40		5.83		4.75
2016	70.85	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	81.20	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	86.40	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_{out} (%)			2.27		4.94		**1.32**		**1.32**		8.71		**1.28**
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) $\frac{d}{dt}x(t) = ax(t) + b_1 t + c$, (iv) $\frac{d}{dt}x(t) = ax(t) + b_1 t + b_2 t^2 + c$, (v) $\frac{d}{dt}x(t) = ax(t) + b_1 t + b_2 t^2 + b_3 t^3 + c$, and (vi) $\frac{d}{dt}y(t) = ay(t) + b_1 t + b_2 t^2 + c$.

are used as a training dataset to construct the 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

The models 1-5 in Table 5 are five integral matching-based ordinary differential equation models with different structures. The only difference among 5 models lies in the forcing terms which are expressed by the linear combination of the time polynomial basis functions. The results obtained from models 1-5 are analyzed in the following.

(i) In comparison with model 5, models 1-4 have lower fitting errors (all the MAPE_{in}s less than 5.0%) and forecasting errors (all the MAPE_{out}s less than 5.0%), and meanwhile, model 5 has more complex structure than models 1-4, indicating that models 1-4 are superior to model 5.

(ii) There are only small differences in the fitting error MAPE_{in}s among models 1-4, whereas models 1, 3 and 4 have much lower forecasting errors in comparison with model 2, indicating that model 2 is inferior to the other three.

(iii) The fitting and forecasting errors of models 3 and 4 are equal to each other, respectively, whereas model 4 has an additional term $b_2 t^2$, indicating that model 3 is more parsimonious (parametrically efficient) and so better according to the Occam's razor principle [74]. In fact, the estimated parameters of models 4 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 model 3 with $\hat{a} = -0.0458$, $\hat{c} = 0.5761$, $\hat{b}_1 = 0.7730$, $\hat{\eta} = 20.8931$.

(iv) Model 3 has a higher fitting error MAPE_{in} but lower forecasting error MAPE_{out} than model 1. The reason for the former is that the estimated initial value of model 3 is much greater than the true value. Without consideration of the first sample, the MAPE_{in}s of models 1 and 3 in the period from 2005 to 2015 are 3.24% and 2.85%, respectively, indicating that model 3 is superior in terms of fitting and forecasting accuracies. In fact, the estimated parameters of model 1 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, model 3 is the best among models 1-5. In addition, we consider the second-order grey polynomial model (model 6 in Table 5 corresponding to model 3) and analyze the similarities and differences. Table 5 and Figure 7(a) show that there are only slight differences, excluding the errors in 2004. Since the fitting values in 2004 are actually the estimated initial values, these differences are the main reason for the differences in performance.

The expressions of models 3 and 6 are $\frac{d}{dt}x_{\mathrm{m}}(t) = -0.04578x_{\mathrm{m}}(t)+0.7730t+0.5761,\ x(1) = 20.8931,\ t \geq 1$ and $\frac{d}{dt}y_{\mathrm{g}}(t) = -0.04578y_{\mathrm{g}}(t) + 0.3865t^2 + 0.9626t + 20.6123,\ y(1) = 21.5509,\ t \geq 1$, respectively; the time response functions are

$$x_{\mathrm{m}}(t) = 16.8847t + 377.1157\exp(-0.04578t) - 356.2318,\ t \geq 1$$

and $y_{\mathrm{g}}(t) = 8.4424t^2 - 347.7895t - 8046.2287\exp(-0.04578t) + 8047.0682,\ t \geq 1$, which corresponds to the restored time response function expressed as

$$x_{\mathrm{g}}(t) = \frac{y(t) - y(t-1)}{t-(t-1)} = 16.8847t + 376.9255\exp(-0.04578t) - 356.2318,\ t \geq 2.$$

By comparing $x_{\mathrm{m}}(t)$ with $x_{\mathrm{g}}(t)$, models 3 and 6 have the same time response functions except for the coefficients of the exponential terms which are caused by the estimated initial values. Overall, model 3 has simpler modelling procedures and is easier to explain, so it is probably the best one to use in this application.

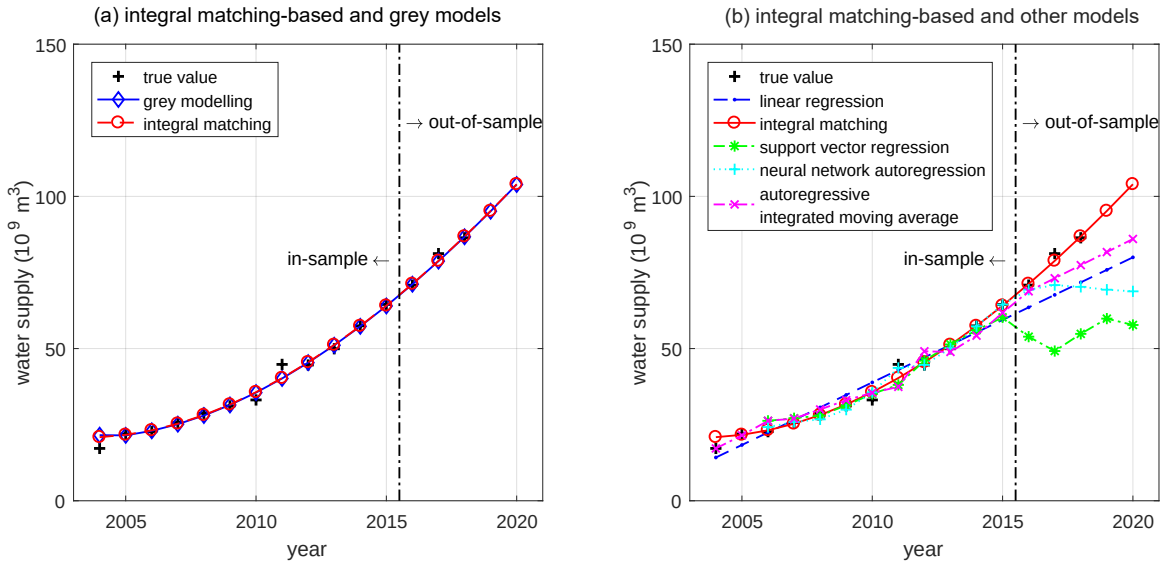

Figure 7: Fitting and forecasting results of other water supply obtained from various models.

6.3. Comparison with other models

The integral matching-based model (model 3) 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 6 and Figure 7(b). ARIMA and NNAR models are implemented by the auto.arima and nnetar functions in forecast package [75], and SVR model by the svm function in e1071 package [76] in R software.

Table 6 shows that all the $\mathrm{MAPE}_{\mathrm{in}}$s of the five models are less than 10% but their $\mathrm{MAPE}_{\mathrm{out}}$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

Table 6: Fitting and forecasting results obtained from integral matching-based and other models.

Year	True	IM [i]		LR [ii]		ARIMA [iii]		NNAR [iv]		SVR [v]	
	value	Value	APE	Value	APE	Value	APE	Value	APE	Value	APE
2004	17.20	20.89	21.47	14.22	17.30	17.19	0.07				
2005	21.96	21.66	1.38	18.33	16.51	21.50	2.09				
2006	22.70	23.14	1.95	22.44	1.13	26.26	15.68	24.21	6.66	26.26	15.70
2007	25.70	25.32	1.49	26.55	3.31	27.00	5.06	25.88	0.71	27.11	5.47
2008	28.74	28.15	2.05	30.66	6.68	30.00	4.38	26.74	6.96	27.98	2.64
2009	31.16	31.61	1.45	34.77	11.58	33.04	6.03	29.93	3.96	31.33	0.55
2010	33.12	35.68	7.72	38.88	17.38	35.46	7.07	35.63	7.57	35.11	6.00
2011	44.80	40.31	10.02	42.99	4.05	37.42	16.47	43.64	2.60	38.07	15.02
2012	44.60	45.50	2.01	47.10	5.59	49.10	10.09	44.63	0.07	46.02	3.18
2013	49.94	51.20	2.53	51.20	2.53	48.90	2.08	50.11	0.34	51.36	2.84
2014	57.46	57.41	0.08	55.31	3.74	54.24	5.60	57.32	0.24	56.05	2.46
2015	64.50	64.10	0.62	59.42	7.87	61.76	4.25	64.46	0.05	60.25	6.59
$MAPE_{in}$ (%)			4.40		8.14		6.57		2.92		6.05
2016	70.85	71.24	0.55	63.53	10.33	68.80	2.89	69.34	2.13	53.91	23.91
2017	81.20	78.82	2.93	67.64	16.70	73.10	9.98	70.85	12.74	49.19	39.42
2018	86.40	86.81	0.48	71.75	16.96	77.40	10.42	70.27	18.67	54.83	36.54
$MAPE_{out}$ (%)			**1.32**		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) IM is the abbreviation of integral matching-based model 3 in Table 5; (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 ϵ-SVR with the embedding dimension equal to 2 and the kernel type being radial basis function.

the autocorrelation in time series and thus results in poor performance. On the contrary, ARIMA consider the autocorrelation, thereby having a better performance than LR. NNAR obtains the lowest 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 has acceptable fitting error ($MAPE_{in} = 4.40\%$) and minimal forecasting error ($MAPE_{out} = 1.32\%$) and so can be considered the best in this example (see Figure 7(b)).

7. Conclusions and outlooks

7.1. Conclusions

In this study, grey and integral matching-based models are both concerned with the data-based modelling in terms of differential equation models: the former estimates the model parameters using Cusum series, while the latter uses the original series. We utilize matrix computation techniques to develop a unified grey model that not only applies to the existing single-variable, multi-variable and multi-output grey forecasting models but can be used also to derive some other novel ones. The integral transformation links the unified grey and integral matching-based models, and the relevant ordinary differential equations prove to be equivalent in a mathematical analysis sense. We show that the Cusum operator is the discrete approximation of the integral transformation formula, so explaining the mechanism of Cusum operator. The integral matching-based model is superior to the grey one due to its simultaneous structural parameter and initial value estimation, as well as its simpler modelling procedures and the easier explanation of its modelling results. The large-scale simulation results show that integral matching-based model out-performs the grey one in terms of the accuracy and robustness to noise; and this is confirmed in the real-world water resources example.

7.2. Outlooks

The results of simulation studies in section 5 show that both grey and integral matching-based models do not perform well when the signal-to-noise ratio is low. The integral matching-based models can be actually rewritten as the state space form

$$\begin{aligned} \text{State equation} \quad & \frac{d}{dt}\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\frac{d}{dt}\boldsymbol{u}(t),\ \boldsymbol{x}(t_1) = \boldsymbol{\eta},\ t \geq t_1 \\ \text{Observation equation} \quad & \tilde{\boldsymbol{x}}(t_k) = \boldsymbol{x}(t_k) + \boldsymbol{e}(k),\ \boldsymbol{e}(k) \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_d) \end{aligned}$$

then similar to the numerical discretization-based equation (23) in integral matching estimation, the pseudo-regression expression of state equation yields the structural model

$$\tilde{\boldsymbol{x}}(t_k) = \boldsymbol{A}\left[\sum_{i=2}^{k}\frac{h_i}{2}\tilde{\boldsymbol{x}}(t_{i-1}) + \sum_{i=2}^{k}\frac{h_i}{2}\tilde{\boldsymbol{x}}(t_i)\right] + \boldsymbol{B}\left[\boldsymbol{u}(t_k) - \boldsymbol{u}(t_1)\right] + \boldsymbol{\eta} + \boldsymbol{\epsilon}(k)$$

where $\boldsymbol{\epsilon}(k)$ is the sum of numerical error and noise error at each item. It is worth noting that the presence of errors-in-variables (also called measurement error in statistics) of the above type induces an asymptotic bias on the parameter estimates which is a function of the signal-to-noise ratio on $\tilde{\boldsymbol{x}}(t_k)$ and is zero only when there is no noise, i.e., $\boldsymbol{\epsilon}(k) = 0, \forall k$ (see section 3.3.2 in [7]).

(i) Other methods of estimation have advantages in this regard, such as the refined instrumental variable estimation [77]; another possibility is two-stage linear least squares [78, 79] where the original time-series observations are de-noised to increase the signal-to-noise ratio at the first stage and then the parameters are estimated using linear east squares at the second stage.

(ii) Both grey and integral matching-based models use linear least squares-based statistical estimation but do not consider the estimation properties of their corresponding estimators. Consequently, future research should be focused on the statistical properties in relation to minimum variance and asymptotic analysis.

(iii) Forecasts obtained from data-based models are inherently uncertain and such forecasts would normally include forecast bounds based on the uncertainty in the data and parameter estimates (see e.g. [7] and the prior references therein). In order to limit the length of the paper, this has not been included but it would be essential in practical applications when using the methods developed in this paper.

(iv) When they are used for forecasting, both grey and integral matching-based models generate the forecasts by solving the associated ordinary differential equations with the structural parameters and initial values set to their estimated values. However, the solutions to ordinary differential equations are sensitive to parameters and initial values, so using a recursive algorithm to estimate the parameters and forecasts, within a Kalman filter setting, is another research direction [7, 80].

Acknowledgment

This work was supported by the National Natural Science Foundation of China (71671090), Joint Research Project of National Natural Science Foundation of China and Royal Society of UK (71811530338), Fundamental Research Funds for the Central Universities of China (NP2018466), China Scholarship Council and Qinglan Project for excellent youth or middle-aged academic leaders in Jiangsu (China).

References

[1] E. S. Gardner Jr, Exponential smoothing: the state of the art, Journal of forecasting 4 (1985) 1–28.
[2] G. E. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, Time series analysis: forecasting and control, John Wiley & Sons, 2015.
[3] P. C. Young, Data-based mechanistic modelling and forecasting globally averaged surface temperature, International Journal of Forecasting 34 (2018) 315–334.
[4] A. J. Smola, B. Schölkopf, A tutorial on support vector regression, Statistics and computing 14 (2004) 199–222.

[5] T. Hill, M. O'Connor, W. Remus, Neural network models for time series forecasts, Management science 42 (1996) 1082–1092.
[6] C. Cheng, A. Sa-Ngasoongsong, O. Beyca, T. Le, H. Yang, Z. Kong, S. T. Bukkapatnam, Time series forecasting for nonlinear and non-stationary processes: a review and comparative study, IIE Transactions 47 (2015) 1053–1071.
[7] P. C. Young, Recursive estimation and time-series analysis: an introduction for the student and practitioner, 2nd ed., Springer Science & Business Media, 2012.
[8] S. L. Brunton, J. L. Proctor, J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, Proceedings of the National Academy of Sciences 113 (2016) 3932–3937.
[9] W.-X. Wang, Y.-C. Lai, C. Grebogi, Data based identification and prediction of nonlinear and complex dynamical systems, Physics Reports 644 (2016) 1–76.
[10] J. Deng, Grey theory and methods in social and economic system (in Chinese), Social Sciences in China (1984) 47–60.
[11] S. Liu, Y. Yang, J. Forrest, Grey data analysis: methods, models and applications, Springer-Verlag, Singapore., 2017.
[12] N. M. Xie, R. Z. Wang, A historic review of grey forecasting models, Journal of Grey System 29 (2017) 1–29.
[13] J. Ne, et al., Invitation to discrete mathematics, Oxford University Press, 2009.
[14] D.-C. Li, C.-J. Chang, W.-C. Chen, C.-C. Chen, An extended grey forecasting model for omnidirectional forecasting considering data gap difference, Applied Mathematical Modelling 35 (2011) 5051–5058.
[15] X. P. Xiao, H. Guo, S. H. Mao, The modeling mechanism, extension and optimization of grey GM(1,1) model, Applied Mathematical Modelling 38 (2014) 1896–1910.
[16] 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, Applied Mathematical Modelling 71 (2019) 543–557.
[17] R. K. Colwell, C. X. Mao, J. Chang, Interpolating, extrapolating, and comparing incidence-based species accumulation curves, Ecology 85 (2004) 2717–2727.
[18] S. Chatterjee, P. Qiu, et al., Distribution-free cumulative sum control charts using bootstrap-based control limits, The Annals of applied statistics 3 (2009) 349–369.
[19] Ø. Borgan, Y. Zhang, Using cumulative sums of martingale residuals for model checking in nested case-control studies, Biometrics 71 (2015) 696–703.
[20] 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 40 (2012) 525–532.
[21] D. C. Li, C. Yeh, C. J. Chang, An improved grey-based approach for early manufacturing data forecasting, Computers & Industrial Engineering 57 (2009) 1161–1167.
[22] 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 40 (2012) 767–773.
[23] 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, Applied Mathematics & Computation 265 (2015) 400–408.
[24] Y. H. Wang, Q. Liu, J. R. Tang, C. W. Bin, L. X. Zhong, Optimization approach of background value and initial item for improving prediction precision of GM(1,1) model, Journal of Systems Engineering & Electronics 25 (2014) 77–82.
[25] C.-S. Shih, Y.-T. Hsu, J. Yeh, P.-C. Lee, Grey number prediction using the grey modification model with progression technique, Applied Mathematical Modelling 35 (2011) 1314–1321.
[26] Y. G. Dang, S. F. Liu, K. J. Chen, The GM models that $x(n)$ be taken as initial value, Kybernetes 33 (2004) 247–254.
[27] Y. H. Wang, Y. G. Dang, Y. Q. Li, S. F. Liu, An approach to increase prediction precision of GM(1,1) model based on optimization of the initial condition, Expert Systems with Applications 37 (2010) 5640–5644.
[28] J. Xu, T. Tan, M. Tu, L. Qi, Improvement of grey models by least squares, Expert Systems with Applications 38 (2011) 13961–13966.
[29] C. I. Chen, S. J. Huang, The necessary and sufficient condition for GM(1,1) grey prediction model, Applied Mathematics & Computation 219 (2013) 6152–6162.
[30] 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, Applied Mathematics & Computation 246 (2014) 648–660.
[31] T. L. Tien, A new grey prediction model FGM(1, 1), Mathematical & Computer Modelling 49 (2009) 1416–1426.
[32] T. X. Yao, S. F. Liu, N. M. Xie, On the properties of small sample of GM(1,1) model, Applied Mathematical Modelling 33 (2009) 1894–1903.
[33] L. F. Wu, S. F. Liu, L. G. Yao, S. L. Yan, The effect of sample size on the grey system model, Applied Mathematical Modelling 37 (2013) 6577–6583.
[34] R. Guo, D. Guo, Random fuzzy variable foundation for grey differential equation modeling, Soft Computing 13 (2009) 185–201.
[35] J. Cui, S. F. Liu, B. Zeng, N. M. Xie, A novel grey forecasting model and its optimization, Applied Mathematical Modelling 37 (2013) 4399–4406.
[36] P.-Y. Chen, H.-M. Yu, Foundation settlement prediction based on a novel NGM model, Mathematical Problems in Engineering 2014 (2014).
[37] W. Y. Qian, Y. G. Dang, S. F. Liu, Grey GM(1,1,t^{α}) model with time power and its application (in Chinese), Systems Engineering – Theory & Practice 32 (2012) 2247–2252.
[38] D. Luo, B. L. Wei, Grey forecasting model with polynomial term and its optimization, Journal of Grey System 29 (2017) 58–69.
[39] X. Ma, Y. S. Hu, Z. B. Liu, A novel kernel regularized nonhomogeneous grey model and its applications, Communications in Nonlinear Science & Numerical Simulation 48 (2016) 51–62.
[40] C.-T. Lin, P.-F. Hsu, B.-Y. Chen, Comparing accuracy of GM(1,1) and grey Verhulst model in Taiwan dental clinics

forecasting, Journal of Grey System 19 (2007) 31–38.
[41] C.-I. Chen, H. L. Chen, S.-P. Chen, Forecasting of foreign exchange rates of taiwan's major trading partners by novel nonlinear grey Bernoulli model NGBM(1,1), Communications in Nonlinear Science and Numerical Simulation 13 (2008) 1194–1204.
[42] T.-L. Tien, A research on the grey prediction model GM(1,n), Applied mathematics & computation 218 (2012) 4903–4916.
[43] B. Zeng, C. Luo, S. Liu, Y. Bai, C. Li, Development of an optimization method for the GM(1, n) model, Engineering Applications of Artificial Intelligence 55 (2016) 353–362.
[44] X. Ma, Z.-b. Liu, The kernel-based nonlinear multivariate grey model, Applied Mathematical Modelling 56 (2018) 217–238.
[45] Z.-X. Wang, D.-J. Ye, Forecasting chinese carbon emissions from fossil energy consumption using non-linear grey multivariable models, Journal of Cleaner Production 142 (2017) 600–612.
[46] P. P. Xiong, Y. Zhang, B. Zeng, T. X. Yao, MGM(1,m) model based on interval grey number sequence and its applications, Grey Systems: Theory and Application 7 (2017) 310–319.
[47] 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, Engineering Applications of Artificial Intelligence 42 (2015) 82–93.
[48] 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, Computers in biology and medicine 105 (2019) 81–91.
[49] L. Wu, S. Liu, Y. Wang, Grey Lotka–Volterra model and its application, Technological Forecasting and Social Change 79 (2012) 1720–1730.
[50] 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, Technological Forecasting and Social Change 106 (2016) 37–44.
[51] B. L. Wei, N. M. Xie, A. Q. Hu, Optimal solution for novel grey polynomial prediction model, Applied Mathematical Modelling 62 (2018) 717–727.
[52] 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 Systems with Applications 41 (2014) 1236–1244.
[53] L.-C. Hsu, A genetic algorithm based nonlinear grey bernoulli model for output forecasting in integrated circuit industry, Expert systems with Applications 37 (2010) 4318–4323.
[54] Z.-X. Wang, An optimized nash nonlinear grey bernoulli model for forecasting the main economic indices of high technology enterprises in china, Computers & Industrial Engineering 64 (2013) 780–787.
[55] J. Lu, W. Xie, H. Zhou, A. Zhang, An optimized nonlinear grey bernoulli model and its applications, Neurocomputing 177 (2016) 206–214.
[56] L. Kong, X. Ma, Comparison study on the nonlinear parameter optimization of nonlinear grey bernoulli model (NGBM(1,1)) between intelligent optimizers, Grey Systems: Theory & Application 8 (2018) 210–226.
[57] H. Guo, X. Xiao, J. Forrest, A research on a comprehensive adaptive grey prediction model CAGM(1,n), Applied Mathematics & Computation 225 (2013) 216–227.
[58] B. Zeng, C. Li, Improved multi-variable grey forecasting model with a dynamic background-value coefficient and its application, Computers & Industrial Engineering 118 (2018) 278–290.
[59] Y.-H. Lin, P.-C. Lee, Novel high-precision grey forecasting model, Automation in construction 16 (2007) 771–777.
[60] Y.-H. Lin, P.-C. Lee, T.-P. Chang, Adaptive and high-precision grey forecasting model, Expert Systems with Applications 36 (2009) 9658–9662.
[61] E. Kayacan, B. Ulutas, O. Kaynak, Grey system theory-based models in time series prediction, Expert systems with applications 37 (2010) 1784–1789.
[62] P. P. Xiong, Y. G. Dang, T. X. Yao, Z. X. Wang, Optimal modeling and forecasting of the energy consumption and production in china, Energy 77 (2014) 623–634.
[63] 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, European Journal of Operational Research 186 (2008) 1184–1200.
[64] P. P. Xiong, S. Huang, M. Peng, X. H. Wu, Examination and prediction of fog and haze pollution using a multi-variable grey model based on interval number sequences, Applied Mathematical Modelling 77 (2020) 1531–1544.
[65] P. Xiong, Y. Dang, X. Wu, X. Li, Combined model based on optimized multi-variable grey model and multiple linear regression, Journal of Systems Engineering and Electronics 22 (2011) 615–620.
[66] I. Dattner, C. A. Klaassen, et al., Optimal rate of direct estimators in systems of ordinary differential equations linear in functions of the parameters, Electronic Journal of Statistics 9 (2015) 1939–1973.
[67] I. Dattner, A model-based initial guess for estimating parameters in systems of ordinary differential equations, Biometrics 71 (2015) 1176–1184.
[68] B. L. Wei, N. M. Xie, L. Yang, Understanding cumulative sum operator in grey prediction model with integral matching, Communications in Nonlinear Science and Numerical Simulation 82 (2020) 105076.
[69] C. F. Van Loan, G. H. Golub, Matrix computations, 4th ed., Johns Hopkins University Press, 2013.
[70] B. L. Wei, N. M. Xie, Y. J. Yang, Data-based structure selection for unified discrete grey prediction model, Expert Systems with Applications 136 (2019) 264–275.
[71] P. C. Young, D. J. Pedregal, W. Tych, Dynamic harmonic regression, Journal of forecasting 18 (1999) 369–394.
[72] 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, International Journal of forecasting 18 (2002) 673–695.
[73] J. S. Famiglietti, The global groundwater crisis, Nature Climate Change 4 (2014) 945.
[74] P. Young, S. Parkinson, M. Lees, Simplicity out of complexity in environmental modelling: Occam's razor revisited, Journal of applied statistics 23 (1996) 165–210.
[75] R. J. Hyndman, Y. Khandakar, Automatic time series forecasting: The forecast package for R, Journal of Statistical

Software 27 (2008) 1548–7660.
[76] E. Dimitriadou, K. Hornik, F. Leisch, D. Meyer, A. Weingessel, M. F. Leisch, Package 'e1071', R Software package (2019).
[77] P. C. Young, Refined instrumental variable estimation: maximum likelihood optimization of a unified box–jenkins model, Automatica 52 (2015) 35–46.
[78] H. Miao, X. Xia, A. S. Perelson, H. Wu, On identifiability of nonlinear ode models and applications in viral dynamics, SIAM review 53 (2011) 3–39.
[79] J. Ramsay, G. Hooker, Dynamic data analysis: modeling data with differential equations, Springer Science & Business Media, 2017.
[80] P. C. Young, Gauss, kalman and advances in recursive parameter estimation, Journal of Forecasting 30 (2011) 104–146.