# On unified framework for nonlinear grey system models: an integro-differential equation perspective

Lu Yang[a], Naiming Xie[a,*], Baolei Wei[a] and Xiaolei Wang[a]

[a]*College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, PR China*

## ARTICLE INFO

## ABSTRACT

Nonlinear grey system models, serving to time series forecasting, are extensively used in diverse areas of science and engineering. However, most research concerns improving classical models and developing novel models, relatively limited attention has been paid to the relationship among diverse models and the modelling mechanism. The current paper proposes a unified framework and reconstructs the unified model from an integro-differential equation perspective. First, we propose a methodological framework that subsumes various nonlinear grey system models as special cases, providing a cumulative sum series-orientated modelling paradigm. Then, by introducing an integral operator, the unified model is reduced to an equivalent integro-differential equation; on this basis, the structural parameters and initial value are estimated simultaneously via the integral matching approach. The modelling procedure comparison further indicates that the integral matching-based integro-differential equation provides a direct modelling paradigm. Next, large-scale Monte Carlo simulations are conducted to compare the finite sample performance, and the results show that the reduced model has higher accuracy and robustness to noise. Applications of forecasting the municipal sewage discharge and water consumption in the Yangtze River Delta of China further illustrate the effectiveness of the reconstructed nonlinear grey models.

*arXiv:2109.03101v2 [stat.AP] 8 Sep 2021*

## 1. Introduction

State space models, a powerful framework for time series, are extensively used to fit the measurement data and then forecast the evolution of dynamic systems. The main feature of state space models is that depicts the underlying process describing dynamic law of the system in terms of states [17]. There exists a diversity of state space models, such as transfer function [26], exponential smoothing [16], and dynamic linear regression [43]. In nature, grey system models proposed by Deng [6] belong to the state space systems. One of the key issues of the grey system models is identifying the governing equations from measurement data for physical understanding, forecasting and controlling. Over the past four decades, a range of grey system models, including linear and nonlinear systems, have been emerging to solve a class of time series forecasting problems. It is worth noting that rather than fitting the original time series, however, grey system models visualize and identify the pattern hidden in the original time series by utilizing cumulative sum (Cusum) operator, distinguishing themselves from other classical state space models. Recently, by employing the integral matching approach, we explained the mechanism of Cusum operator from the perspectives of mathematical analysis and parameter estimations [36] . Then, on this basis, a unified modelling paradigm for linear grey forecasting models is proposed to link grey system models with dynamic data analysis [34, 35].

Nonlinear grey system models, aimed at describing more complex systems, have broader applicability than linear ones, having attracted considerable attention in the grey system community. Up to now, most

research is focused on improving the classical models and developing novel ones, which enriches nonlinear grey system model families. As an emerging method, however, there remain several challenges needed to be addressed. First, there exist many nonlinear grey system models with different forms, making it difficult for researchers to perform property analysis. Hence it is required to form a unified methodological framework, fully illustrating the modelling paradigm. The distinguishing feature of nonlinear grey models is fitting the Cusum series, but the reason why nonlinear grey systems models use the Cusum operator has not been well explained. So, is it possible that fitting the original time series instead of Cusum one but achieving the same forecasts? If so, the mechanism of nonlinear grey models will be revealed, making the modelling results easier to understand and explain. Besides, both parameter estimation and initial value selection affect the forecasting results. Current research tends to use a two-step method: estimating the structural parameters and subsequently selecting the initial condition via a given strategy (see [34] for the most used three strategies). The inconsistent objective functions in two separate stages may introduce extra errors and degrade outcomes significantly [37], so it is necessary to develop a one-step parameter estimation method.

Therefore, in this work, we seek a unified framework for nonlinear grey system models and reconstruct the resultant unified form via an integro-differential equation. The principal contributions are summarized as follows:

(1) We propose a new methodological framework for nonlinear grey system models, which not only has the ability to unify the existing systems, but may also induce novel ones, providing the basis of nonlinear grey modelling paradigm, i.e., the Cusum series-orientated nonlinear differential equation models.

(2) By introducing an integral operator, we reduce the unified representation to an integro-differential equation, allowing us to fit the original time series directly and analyse the mechanism of Cusum operator from both mathematical analysis and parameter estimation perspectives.

(3) Based on the reduced integro-differential equation, a new one-step parameter estimation approach, integral matching, is introduced to estimate structural parameters and initial value simultaneously.

The remaining parts are organized as follows. Section 2 reviews the existing nonlinear grey system models from the original to the extended ones. Section 3 presents the unified framework for nonlinear grey system models. Section 4 proposes the reduced reconstruction via an integro-differential equation. Section 5 conducts Monte Carlo simulations to evaluate the finite sample performance. Section 6 provides a real-world application and section 7 concludes the work.

## 2. Literature review

In this section, we review the existing nonlinear grey system models from two viewpoints: the basic nonlinear grey model and the extended ones.

### 2.1. The basic grey Verhulst model

The grey Verhulst model (GVM(1,1)) [6], where the first "1" denotes the order of derivative and the second "1" denotes the dimensional of variables, aims at fitting inverted U-shaped time series and lays the foundation for the development of nonlinear grey system models.

**Definition 1.** [34] *For a time series with n samples $X(t) = \{x(t_1), x(t_2), \cdots, x(t_n)\}$, the Cusum series is defined as $Y(t) = \{y(t_1), y(t_2), \cdots, y(t_n)\}$ for $y(t_k) = \sum_{i=1}^{k} h_i x(t_k)$, where $h_1 = 1$ and $h_k = t_k - t_{k-1}$ for $k \geq 2$.*

The grey Verhulst model consists of the differential equation

$$\frac{d}{dt}y(t) = ay(t) + b\,(y(t))^2 , \ t \geq t_1 \tag{1}$$

and the corresponding discrete-time equation

$$x(t_k) = a\,[\lambda y(t_{k-1}) + (1-\lambda)y(t_k)] + b\,[\lambda y(t_{k-1}) + (1-\lambda)y(t_k)]^2 , \ k = 2,\,3\cdots,\,n \tag{2}$$

where $\lambda \in [0,1]$ is referred to as a background coefficient whose value is always set to 0.5.

The structural parameters can be easily estimated by the least squares criterion

$$[\hat{a}\ \hat{b}]^\top = \underset{a,b}{\arg\min} \left\| \mathbf{Y} - \mathbf{B}\,[a\ b]^\top \right\|_2^2 = \left(\mathbf{B}^\top \mathbf{B}\right)^{-1} \mathbf{B}^\top \mathbf{Y}$$

where

$$\mathbf{B} = \begin{bmatrix} \frac{y(t_1)+y(t_2)}{2} & \left(\frac{y(t_1)+y(t_2)}{2}\right)^2 \\ \frac{y(t_2)+y(t_3)}{2} & \left(\frac{y(t_2)+y(t_3)}{2}\right)^2 \\ \vdots & \vdots \\ \frac{y(t_{n-1})+y(t_n)}{2} & \left(\frac{y(t_{n-1})+y(t_n)}{2}\right)^2 \end{bmatrix}, \ \mathbf{Y} = \begin{bmatrix} x(t_2) \\ x(t_3) \\ \vdots \\ x(t_n) \end{bmatrix}.$$

Let the initial value be $y(t_1) = \eta_y$, the closed form solution of equation (1) is

$$\hat{y}(t) = \left[ -\frac{\hat{b}}{\hat{a}} + e^{-\hat{a}(t-t_0)} \cdot \left( \frac{1}{\eta_y} + \frac{\hat{b}}{\hat{a}} \right) \right]^{-1}$$

then the desired forecasts $\hat{X}$ can be calculated by the inverse Cusum operator derived from Definition 1, that is,

$$\begin{cases} \hat{x}(t_1) = \eta_y \\ \hat{x}(t_k) = \frac{1}{h_k}\left(\hat{y}(t_k) - \hat{y}(t_{k-1})\right), \ k = 2,\,3,\cdots,\,n+r \end{cases}$$

where $r$ is the forecasting horizon.

Subsequently, a great deal of effort has been devoted to improve the accuracy and the existing research can be divided into the following three types: (i) improving the structural parameter estimates via intelligence algorithm [24, 31], weighted least squares [27], and parameter transformation [10], (ii) optimizing the initial value selection strategy [3], and (iii) searching the optimal background coefficient [8].

## 2.2. Extended nonlinear grey system models

Following the similar ideas and modelling procedures, the extended models utilize other nonlinear differential equations to fit Cusum series. A rough classification of the nonlinear grey system models includes two categories: single-output and multi-output models, and we present the main extensions in Table 1.

Table 1 shows that single-output models subsume single-variable and multi-variable models. Single-variable models share a similar representation. For instance, when $c = 0$, GRBM(1,1) can reduce to NBGM(1,1) and, then, if $\gamma$ in NGBM(1,1) equals 2, then GVM(1,1) is obtained. For multi-variable exten-

Table 1: The coupled equations of continuous-time nonlinear grey system models

| Type | Model | Differential equation | Difference equation | Ref |
|---|---|---|---|---|
| single-output | GVM(1,1) | $\frac{d}{dt}y(t) = ay(t) + b\left(y(t)\right)^2$ | $x(t_k) = a\left(z(t_k)\right) + b\left(z(t_k)\right)^2$ | [24] |
| | GGVM(1,1) | $\frac{d}{dt}y(t) = ay(t) + b\left(y(t)\right)^2 + c$ | $x(t_k) = a\left(z(t_k)\right) + b\left(z(t_k)\right)^2 + c$ | [46] |
| | GRM(1,1) | $\frac{d}{dt}y(t) = ay(t) + b\left(y(t)\right)^2 + c(k-1)^{\gamma} + d$ | $x(t_k) = az(t_k) + b\left(z(t_k)\right)^2 + c(k-1)^{\gamma} + d$ | [12] |
| | NGM(1,1,$\alpha$) | $\frac{d}{dt}y(t) = a\left(y(t)\right)^{\alpha} + b$ | $x(t_k) = a\left(z(t_k)\right)^{\alpha} + b$ | [29] |
| | NGBM(1,1) | $\frac{d}{dt}y(t) = ay(t) + b\left(y(t)\right)^{\gamma}$ | $x(t_k) = a\left(z(t_k)\right) + b\left(z(t_k)\right)^{\gamma}$ | [2] |
| | GRBM(1,1) | $\frac{d}{dt}y(t) = ay(t) + b\left(y(t)\right)^{\gamma} + c$ | $x(t_k) = a\left(z(t_k)\right) + b\left(z(t_k)\right)^{\gamma} + c$ | [41] |
| | NGBMC(1,n) | $\frac{d}{dt}y_1(rp+t) = ay_1(rp+t) + \left(\sum\limits_{i=2}^{n} b_i y_i(t) + u\right)\left(y_1(rp+t)\right)^2$ | no explicit from | [44] |
| | KRNGM(1,n) | $\frac{d}{dt}y_{\text{out}}(t) = ay_{\text{out}}(t) + \boldsymbol{\omega}^{\top}\boldsymbol{\phi}(y_{\text{in}}(t)) + u$ | $x_{\text{out}}(t_k) = az_{\text{out}}(t_k) + \boldsymbol{\omega}^{\top}\boldsymbol{\phi}(y_{\text{in}}(t_k)) + u$ | [18] |
| multi-output | GLVM(1,2) | no explicit form | $\begin{cases} x_1(t_k) = a_1 z_1(t_k) + b_1\left(z_1(t_k)\right)^2 + c_1 z_1(t_k)z_2(t_k) \\ x_2(t_k) = a_2 z_2(t_k) + b_2\left(z_2(t_k)\right)^2 + c_2 z_1(t_k)z_2(t_k) \end{cases}$ | [13] |

[1] $z_i(t_k) = \omega y_i(t_k) + (1-\omega)y_i(t_{k-1})$

sions, in particular, KRNGM(1,n) uses kernel tricks to introduce other variables. Further ones lead to grey output models, but to our knowledge, research of multi-output models is sparse and mainly focused on the grey prey-predator system (GLVM(1,2)) concerning accuracy improvement [38, 19] and stability analysis [13, 32].

Another extension is hybrid models, including the grey time-delayed Verhulst model which employs a time-delay differential equation [30], the F-NGBM(1,1) model which uses the Fourier series as residual correction tool [4], and the metabolic nonlinear grey–autoregressive integrated moving average model which couples NGM(1,1,$\alpha$) and ARIMA models [28].

Similar to the research route of grey Verhulst model, a range of studies were implemented to optimize the aforementioned models, that is, the structural parameter, initial value and background coefficient optimization for the single-variable cases [15, 45], the multi-variable cases [9], and the multi-output cases [14].

## 3. Unified framework for nonlinear grey system models

In this section, we propose a unified framework for existing nonlinear grey system models. For a $d$-dimensional state vector $\boldsymbol{x}(t)$, the observations are sampled at time points $\{t_1, t_2, \cdots, t_n\}$ and arranged into the following matrix:

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

Then, the corresponding Cusum operator matrix is defined as

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

where $y(t_i) = \sum_{i=1}^{k} h_i x(t_i)$ for $h_1 = 1$ and $h_k = t_k - t_{k-1}$, $k \geq 2$.

Nonlinear grey system models utilize first order nonlinear differential equations to describe the evolution of the Cusum variable. Consider a unified representation

$$\frac{d}{dt} \boldsymbol{y}(t) = \boldsymbol{\theta}_{\mathrm{L}} \boldsymbol{y}(t) + \boldsymbol{\theta}_{\mathrm{N}} \boldsymbol{N}(\boldsymbol{y}(t)) + \boldsymbol{\beta}, \ \boldsymbol{y}(t_1) = \boldsymbol{\eta}_y, \ t \geq t_1 \tag{3}$$

where $\boldsymbol{y}(t) \in \mathbb{R}^d$ is the Cusum state, $\boldsymbol{N}(\boldsymbol{y}(t)) : \mathbb{R}^d \to \mathbb{R}^p$ is a $p$-dimensional nonlinear vector function, $\boldsymbol{\eta}_y$ is the unknown initial value, $\boldsymbol{\theta}_{\mathrm{L}} \in \mathbb{R}^{d \times d}$, $\boldsymbol{\theta}_{\mathrm{N}} \in \mathbb{R}^{d \times p}$ and $\boldsymbol{\beta} \in \mathbb{R}^d$ are the unknown structural parameters.

Note that, here, we explicitly indicate that the dynamics (3) have both linear and nonlinear contributions for the convenience of parameter estimation. Then, the two-step least squares is utilized to estimate the structural parameters and initial value in succession. In the first step, using the implicit Midpoint method gives the corresponding discrete-time equation

$$\frac{\boldsymbol{y}(t_k) - \boldsymbol{y}(t_{k-1})}{t_k - t_{k-1}} = \boldsymbol{x}(t_k) \approx \boldsymbol{\theta}_{\mathrm{L}} \frac{\boldsymbol{y}(t_k) + \boldsymbol{y}(t_{k-1})}{2} + \boldsymbol{\theta}_{\mathrm{N}} \boldsymbol{N}\left(\frac{\boldsymbol{y}(t_k) + \boldsymbol{y}(t_{k-1})}{2}\right) + \boldsymbol{\beta} + \boldsymbol{\epsilon}(k) \tag{4}$$

where $\boldsymbol{\epsilon}(k)$ is the model error and $\frac{1}{2}(\boldsymbol{y}(t_k) + \boldsymbol{y}(t_{k-1}))$ is referred to as background value in the grey system terminology. By substituting $k = 2, 3, \cdots, n$ into equation (4) and arranging the resulting $n-1$ algebraic equations into a matrix form, we have

$$\boldsymbol{X} = \Theta(\boldsymbol{y})\Xi + \Gamma \tag{5}$$

where

$$\Xi = \begin{bmatrix} \boldsymbol{\theta}_{\mathrm{L}}^\top \\ \boldsymbol{\theta}_{\mathrm{N}}^\top \\ \boldsymbol{\beta}^\top \end{bmatrix}, \ \boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}^\top(t_2) \\ \boldsymbol{x}^\top(t_3) \\ \vdots \\ \boldsymbol{x}^\top(t_n) \end{bmatrix}, \ \Theta(\boldsymbol{y}) = \begin{bmatrix} \frac{\boldsymbol{y}^\top(t_1) + \boldsymbol{y}^\top(t_2)}{2} & \boldsymbol{N}^\top\left(\frac{\boldsymbol{y}(t_1) + \boldsymbol{y}(t_2)}{2}\right) & 1 \\ \frac{\boldsymbol{y}^\top(t_2) + \boldsymbol{y}^\top(t_3)}{2} & \boldsymbol{N}^\top\left(\frac{\boldsymbol{y}(t_2) + \boldsymbol{y}(t_3)}{2}\right) & 1 \\ \vdots & \vdots & \vdots \\ \frac{\boldsymbol{y}^\top(t_{n-1}) + \boldsymbol{y}^\top(t_n)}{2} & \boldsymbol{N}^\top\left(\frac{\boldsymbol{y}(t_{n-1}) + \boldsymbol{y}(t_n)}{2}\right) & 1 \end{bmatrix}, \ \Gamma = \begin{bmatrix} \boldsymbol{\epsilon}^\top(2) \\ \boldsymbol{\epsilon}^\top(3) \\ \vdots \\ \boldsymbol{\epsilon}^\top(n) \end{bmatrix}.$$

This then allows for the formulation of a regression problem to estimate the structural parameters:

$$\min_{\Xi} \ \mathcal{L}(\Xi) = \left\| \boldsymbol{X} - \Theta(\boldsymbol{y})\Xi \right\|_{\mathsf{F}}^2 \tag{6}$$

where $\| \cdot \|_{\mathsf{F}}$ is the Frobenius norm. Note, problem (6) is a linear least squares problem due to the separable parameters. Differentiating $\mathcal{L}(\Xi)$ with respect to $\Xi$ yields

$$\frac{\partial}{\partial \Xi} \mathcal{L}(\Xi) = \frac{\partial}{\partial \Xi} \mathrm{Tr}\left([\boldsymbol{X} - \Theta(\boldsymbol{y})\Xi]^\top [\boldsymbol{X} - \Theta(\boldsymbol{y})\Xi]\right) = -2\Theta^\top(\boldsymbol{y})\Xi + 2\Theta^\top(\boldsymbol{y})\Theta(\boldsymbol{y})\Xi$$

where, setting the derivative to zero gives the least-squares estimates

$$\hat{\Xi} = \left(\Theta^\top(\boldsymbol{y})\Theta(\boldsymbol{y})\right)^{-1} \Theta^\top(\boldsymbol{y})\boldsymbol{X}.$$

In the second step, by substituting the estimated structural parameters into equation (3) and solving the resultant differential equations, we have the solution (also termed as time response function) expressed as

$$\hat{\boldsymbol{y}}(t) = \boldsymbol{F}(\boldsymbol{\eta}_y, \hat{\Xi}; t) \tag{7}$$

where $\hat{y}(t_1) = \eta_y \in \mathbb{R}^d$ is unknown initial condition. Note, the closed form solutions of equation (3) are often not available due to the nonlinearity. We can alternatively calculate the numerical solutions via numerical simulation scheme such as the Runge-Kutta algorithm [20]. It is obvious that $\eta_y$ is crucial to the solution (7), and three popular initial condition selection strategies are as follows:

(1) the fixing first point strategy: $\hat{\eta}_y$ is the solution to $y(t_1) = \hat{y}(t_1) = F(\eta_y, \hat{\Xi}; t_1)$;

(2) the fixing last point strategy: $\hat{\eta}_y$ is the solution to $y(t_n) = \hat{y}(t_n) = F(\eta_y, \hat{\Xi}; t_n)$;

(3) the residual error correction strategy: $\hat{\eta}_y = \arg\min_{\eta_y} \left\{ \sum_{i=1}^{n} \left\| F(\eta_y, \hat{\Xi}; t_i) - y(t_i) \right\|_2^2 \right\}$.

Finally, substituting time points $\{t_i\}_{i=1}^{n+r}$ into equation (7) gives the fitting and forecasting values $\hat{y}(t)$ of Cusum series and subsequently, by applying the inverse Cusum operator, the forecasts corresponding to the original time series are given through

$$\hat{x}(t_k) = \begin{cases} \hat{y}(t_1), & k = 1, \\ \frac{1}{h_k} \left( \hat{y}(t_k) - \hat{y}(t_{k-1}) \right) & k = 2, 3, \cdots, r. \end{cases}$$

To conclude, we present a unified representation for nonlinear grey system models and in the following, we further discuss the backwards compatibility of this unified, including the single-output and multi-output grey nonlinear models.

**Remark 1.** *Let the dimension of the system be $d = 1$. The unified model (3) yields multiple families of single-output nonlinear grey system models with different $N(y)$.*

(1) *Considering that $N(y)$ consists of polynomial term, the differential equation and the corresponding discrete-time equation can be written as*

$$\frac{d}{dt} y(t) = a y(t) + \sum_{l=1}^{p+1} b_l \left( y(t) \right)^l + \beta$$

*and*

$$x(t_k) = a \frac{y(t_{k-1}) + y(t_k)}{2} + \sum_{l=1}^{p+1} b_l \left( \frac{y(t_{k-1}) + y(t_k)}{2} \right)^l + \beta + \epsilon(t_k).$$

(2) *Considering that $N(y)$ consists of power form, the differential equation and the corresponding discrete-time equation can be written as*

$$\frac{d}{dt} y(t) = a y(t) + b \left( y(t) \right)^\gamma + \beta$$

*and*

$$x(t_k) = a \frac{y(t_{k-1}) + y(t_k)}{2} + b \left( \frac{y(t_{k-1}) + y(t_k)}{2} \right)^\gamma + \beta + \epsilon(t_k).$$

Remark 1 demonstrates that the unified model subsumes a number of nonlinear grey systems model families, although we only give two scenarios here. Furthermore, the polynomial and power families cover all the single-variable models in Table 1. In practice, particularly, the power model takes $\gamma$ as a hyperparameter which can be determined by optimization methods, such as particle swarm algorithm [7] and line search [42].

**Remark 2.** *Let the dimension of the state vector be $d \geq 2$. Similar to the extensions in Remark 1, if $\mathbf{N}(\mathbf{y})$ consists of polynomial term $\begin{bmatrix} \mathbf{y}^{P_2} & \mathbf{y}^{P_3} & \cdots \end{bmatrix}$, the unified model yields a number of multi-output nonlinear grey models. Here, higher polynomials are denoted as $\mathbf{y}^{P_2}$, $\mathbf{y}^{P_3}$, etc., where $\mathbf{y}^{P_2}$ denotes the quadratic nonlinearities in $\mathbf{y}$: $\begin{bmatrix} y_1^2 & y_1 y_2 & \cdots & y_1 y_d & y_2^2 & \cdots & y_d^2 \end{bmatrix}$.*

*Among them is the most simple form having the continuous- and corresponding discrete-time equation expressed as*

$$
\begin{cases}
\frac{d}{dt} y_1(t) = y_1(t) \left( a_{11} + \sum_{i=1}^{d} b_{1i} y_i(t) \right) \\
\frac{d}{dt} y_2(t) = y_2(t) \left( a_{21} + \sum_{i=1}^{d} b_{2i} y_i(t) \right) \\
\vdots \\
\frac{d}{dt} y_d(t) = y_d(t) \left( a_{d1} + \sum_{i=1}^{d} b_{di} y_i(t) \right)
\end{cases}
\quad \text{and} \quad
\begin{cases}
x_1(t_k) = z_1(t) \left( a_{11} + \sum_{i=1}^{d} b_{1i} z_i(t) \right) \\
x_2(t_k) = z_2(t) \left( a_{21} + \sum_{i=1}^{d} b_{2i} z_i(t) \right) \\
\vdots \\
x_d(t_k) = z_d(t) \left( a_{d1} + \sum_{i=1}^{d} b_{di} z_i(t) \right)
\end{cases}
$$

*where $z_i(t_k) = \frac{1}{2} y_i(t_{k-1}) + \frac{1}{2} y_i(t_k)$, $i = 1, 2, \cdots, d$.*

Remark 2 shows that the unified model can yield the high-dimensional system that are employed to describe interactions between species and, additionally, if $d = 2$, then the unified model is actually the classical grey Lotka-Volterra model [13]. Furthermore, similar to the expansions in Remark 1, the unified model could deduce some other novel multi-output models.

**Remark 3.** *If there exist forcing terms, then the unified model can be extended to*

$$
\frac{d}{dt} \mathbf{y}(t) = \boldsymbol{\theta}_L \mathbf{y}(t) + \boldsymbol{\theta}_N \mathbf{N}(\mathbf{y}(t)) + \boldsymbol{\theta}_F \mathbf{u}(t) + \boldsymbol{\beta}, \ \mathbf{y}(t_1) = \boldsymbol{\eta}_y \tag{8}
$$

*where $\mathbf{u}(t) \in \mathbb{R}^{\ell}$ is a known input vector which is independent of $\mathbf{y}(t)$, $\boldsymbol{\eta}_y$ is unknown initial value, $\boldsymbol{\theta}_L \in \mathbb{R}^{d \times d}$, $\boldsymbol{\theta}_N \in \mathbb{R}^{d \times p}$, $\boldsymbol{\theta}_F \in \mathbb{R}^{d \times \ell}$ and $\boldsymbol{\beta} \in \mathbb{R}^d$ are structural parameters.*

*In light of equation (4), we obtain the corresponding discrete-time equation*

$$
\mathbf{x}(t_k) \approx \boldsymbol{\theta}_L \frac{\mathbf{y}(t_k) + \mathbf{y}(t_{k-1})}{2} + \boldsymbol{\theta}_N \mathbf{N} \left( \frac{\mathbf{y}(t_k) + \mathbf{y}(t_{k-1})}{2} \right) + \boldsymbol{\theta}_F \frac{\mathbf{u}(t_k) + \mathbf{u}(t_{k-1})}{2} + \boldsymbol{\beta} + \boldsymbol{\epsilon}(k).
$$

Remark 3 gives an extension principle of the unified nonlinear grey system model when introducing forcing terms to nonlinear grey models. For instance, considering the single-output scenario, if $\mathbf{u}(t)$ consists of other state variables, then we can obtain a number of multi-variable models, including the NGBMC(1,n) and KRGBM(1,n) in Table 1.

Remarks 1-3 show that the unified model has the capacity to represent classical grey nonlinear models, including but not limited to the existing single-output, and multi-output models. The extension of the above parameter estimation procedures to equation (8) is straightforward and tedious. In order to be focused, we only discuss the model (3) in the following.

## 4. Reconstruction of nonlinear grey models with an integro-differential equation

In this section, we simplify the unified nonlinear grey model into a reduced-order integro-differential equation; then, by using the integral matching approach which consists of an integral operator and the least squares [5], we estimate the structural parameters and initial conditions simultaneously; further, we

compare the modelling procedures of nonlinear grey models with those of the integro-differential equation-based ones.

## 4.1. Reduced-order integro-differential equation model

**Theorem 1.** *Let*

$$y(t) = \eta_y + \int_{t_1}^{t} x(\tau)d\tau, \ t \geq t_1 \tag{9}$$

*where $\eta_y \in \mathbb{R}^d$ is a real vector, the aforementioned unified representation (3) is equivalent to an integro-differential equation expressed as*

$$\frac{d}{dt}x(t) = \theta_L x(t) + \theta_N x(t)\frac{d}{dt}N\left(\eta_y + \int_{t_1}^{t} x(\tau)d\tau\right), \ x(t_1) = \eta_x, \ t \geq t_1 \tag{10}$$

*where $\eta_x = \theta_L \eta_y + \theta_N N(\eta_y) + \beta$.*

*Proof.* To begin with the necessity, that is, equation (3) can be reduced to equation (10). Substituting equation (9) into equation (3) gives

$$x(t) = \theta_L x(t) + \theta_N N\left(\eta_y + \int_{t_1}^{t} x(\tau)d\tau\right) + \beta$$

where, differentiating both sides with respect to $t$ and using the chain rule gives equation (10). Then, by combining equations (3) and (9), the initial value can be obtained as

$$\eta_x = \frac{d}{dt}y(t)\Big|_{t=1} = \theta_L \eta_y + \theta_N N(\eta_y) + \beta.$$

Conversely, integrating both sides of equation (10) with respect $t$, we have, in light of equation (9),

$$\int_{t_1}^{t} dx(\tau) = \theta_L \int_{t_1}^{t} x(\tau)d\tau + \theta_N \int_{t_1}^{t} x(\tau)\frac{d}{d\tau}N\left(\eta_y + \int_{t_1}^{\tau} x(s)ds\right)d\tau.$$

Manipulating the second term at the right-hand side through the method of integration by substitution yields

$$\int_{t_1}^{t} dx(\tau) = \theta_L \int_{t_1}^{t} x(\tau)d\tau + \theta_N \int_{t_1}^{t} \frac{d}{d\tau}N\left(y(\tau)\right)dy(\tau) = \theta_L \int_{t_1}^{t} x(\tau)d\tau + \theta_N \int_{t_1}^{t} dN\left(y(\tau)\right)$$

then, using the Newton-Leibniz formula leads to

$$\begin{aligned} x(t) &= \theta_L\left(y(t) - \eta_y\right) + \theta_N\left(N(y(t)) - N(\eta_y)\right) + \eta_x \\ &= \theta_L y(t) + \theta_N N(y(t)) + \eta_x - \theta_L \eta_y - \theta_N N(\eta_y) \end{aligned} \tag{11}$$

where, particularly, the left-hand side is

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

and substituting the initial condition $\eta_x = \theta_L \eta_y + \theta_N N(\eta_y) + \beta$ into equation (11), we obtain the desired equation (3). □

Theorem 1 shows that for a given nonlinear grey equation (3), we can always find an equivalent

reduced-order integro-differential equation (10). Conversely, equation (10) can be integrated into the unified model (3) with the initial value $\eta_y$ satisfying $\eta_x = \theta_L \eta_y + \theta_N N(\eta_y) + \beta$. Note that the constant vector $\beta$ provides $d$ degrees of freedom corresponding the initial value $\eta_y$. Let us illustrate it with an example.

**Example 1.** *Supposing a single-output nonlinear grey model (the GGVM(1,1) model) is*

$$\frac{d}{dt}y(t) = ay(t) + b\left(y(t)\right)^2, \; y(t_1) = \eta_y, \; t \geq 0 \tag{12}$$

*then the equivalent reduced-order integro-differential equation is*

$$\frac{d}{dt}x(t) = ax(t) + 2bx(t)\left(\eta_y + \int_{t_1}^{t} x(\tau)d\tau\right), \; x(t_1) = \eta_x, \; t \geq 0. \tag{13}$$

*where the initial value satisfy $\eta_x = a\eta_y + b\eta_y{}^2$. It is easy to verify that their closed form solutions satisfy equation (9), that is, the closed form solutions of equations (12) and (13) are*

$$y(t) = \left[-\frac{b}{a} + e^{-a(t-t_1)} \cdot \left(\frac{1}{\eta_y} + \frac{b}{a}\right)\right]^{-1} \; and \; x(t) = \frac{e^{-a(t-t_1)}\left(\frac{b}{a} + \frac{1}{\eta}\right)}{\left[\frac{b}{a} - e^{-a(t-t_1)}\left(\frac{b}{a} + \frac{1}{\eta}\right)\right]^2}, \; respectively.$$

Specially, if $\eta_x = \eta_y = \eta$, the traditional Cusum operator is the left rectangle approximation of equation (9), which has been proved in the literature [35, 42]. In this sense, the integral operator can be regarded as a continuous-time generalisation of the Cusum operator. In such a case, for a given integro-differential equation (10), we can always find an equivalent nonlinear grey system model (3) with initial value $y(t_1) = \eta_x$.

## 4.2. Integral matching for estimating parameters and initial values simultaneously

The reduced-order integro-differential equation concerns the state variable rather than the Cusum state, making the direct modelling possible. By using a state-space form, the reconstructed model can be expressed as

$$\text{Observation equation } x(t_k) = s(t_k) + e(k), k = 1, 2, \cdots, n \tag{14}$$

$$\text{State equation } \frac{d}{dt}s(t) = \theta_L s(t) + \theta_N s(t)\frac{d}{dt}N\left(\zeta + \int_{t_1}^{t} s(\tau)d\tau\right), \; s(t_1) = \eta, \; t \geq t_1 \tag{15}$$

where $x(t_k)$ is the observation, $e(k)$ is the measurement noise, $s(t) \in \mathbb{R}^d$ is the state variable, $N(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^p$ is a $p$-dimensional nonlinear vector function of the integral operator, $\theta_L \in \mathbb{R}^{d \times d}$ is the unknown parameters of linear term, $\theta_N \in \mathbb{R}^{d \times p}$ is unknown structural parameter of nonlinear term, $\zeta \in \mathbb{R}^d$ is an unknown constant parameter, and $\eta \in \mathbb{R}^d$ is the unknown initial condition.

It is obvious that equation (15) is nonlinear in $\zeta$, so that this parameter estimation problem falls in the general class of nonlinear least square problem (NLS). Recall that Theorem 1 supports the flexibility selection of $\zeta$. Therefore, for ease of calculation, we take $\zeta = \eta$, leading to a new form

$$\frac{d}{dt}s(t) = \theta_L s(t) + \theta_N s(t)\frac{d}{dt}N\left(\eta + \int_{t_1}^{t} s(\tau)d\tau\right), \; s(t_1) = \eta, \; t \geq t_1. \tag{16}$$

In what will follow, the integral matching approach (also called the direct integral method), is utilized to estimate the unknown structural parameters and initial value simultaneously.

Integrating equation (16) with respect $t$ over the interval $[t_1, t_k]$, and combining Theorem 1, we have

$$s(t_k) = \theta_{\mathrm{L}} \int_{t_1}^{t_k} s(\tau)d\tau + \theta_{\mathrm{N}} \left[ N\left(\eta + \int_{t_1}^{t_k} s(\tau)d\tau\right) - N(\eta) \right] + \eta \tag{17}$$

where, by using piecewise linear integration formula, the integral term $\int_{t_1}^{t_k} s(\tau)d\tau$ is approximated as

$$\tilde{s}(t_k) \approx \frac{1}{2}\sum_{i=2}^{k} h_i s(t_{i-1}) + \frac{1}{2}\sum_{i=2}^{k} h_i s(t_i).$$

Due to the state variable $s(t)$ is not available, we use the noisy observation $x(t)$ instead, then

$$\tilde{x}(t_k) \approx \frac{1}{2}\sum_{i=2}^{k} h_i x(t_{i-1}) + \frac{1}{2}\sum_{i=2}^{k} h_i x(t_i). \tag{18}$$

Correspondingly, the pseudo-nonlinear regression problem becomes

$$x(t_k) = \theta_{\mathrm{L}} \tilde{x}(t_k) + \theta_{\mathrm{N}}\left[N(\eta + \tilde{x}(t_k)) - N(\eta)\right] + \eta + e(t_k) \tag{19}$$

where $e(t_k)$ is the sum of discretization error and measurement noise. Then, we utilize the change of basis to show how to derive this nonlienar least squares problem to a linear version.

**Lemma 1.** *Suppose that* $N(\cdot)$ *consists of polynomial terms. Depending on whether there exist interaction terms or not we get the following alternatives.*

1. *If* $d = 1$ *and* $N(x) = \begin{bmatrix} x^2, & x^3, & \cdots, & x^{p+1} \end{bmatrix}^\top$, *then*

$$\theta_L \tilde{x}(t_k) + \theta_N \left[N(\eta + \tilde{x}(t_k)) - N(\eta)\right] + \eta = \left(\theta_L + \theta_L \phi\right)\tilde{x}(t_k) + \left(\theta_N \varphi\right)N(\tilde{x}(t_k)) + \eta \tag{20}$$

   *where*

$$\phi = \begin{bmatrix} \binom{2}{1}\eta^1 \\ \vdots \\ \binom{p+1}{1}\eta^p \end{bmatrix}, \quad \varphi = \begin{bmatrix} \binom{2}{2}\eta^0 & & \\ \vdots & \ddots & \\ \binom{p+1}{2}\eta^{p-1} & \cdots & \binom{p+1}{p+1}\eta^0 \end{bmatrix}.$$

2. *If* $d \geq 2$ *and* $N(x) = \begin{bmatrix} x_1^2, & x_1 x_2, & \cdots, & x_1 x_d, & x_2^2, & \cdots, & x_d^2 \end{bmatrix}^\top$, *then*

$$\theta_L \tilde{x}(t_k) + \theta_N \left[N(\eta + \tilde{x}(t_k)) - N(\eta)\right] + \eta = \left(\theta_L I_d + \theta_N \psi\right)\tilde{x}(t_k) + \theta_N N(\tilde{x}(t_k)) + \eta \tag{21}$$

   *where* $\psi = \begin{bmatrix} \psi_d^\top & \psi_{d-1}^\top & \cdots & \psi_1^\top \end{bmatrix}^\top$ *for*

$$\psi_{d-i} = \begin{bmatrix} 0 & \cdots & 0 & 2\eta_{i+1} & & & 0 \\ 0 & \cdots & 0 & \eta_{i+2} & \eta_{i+1} & & \\ \vdots & \ddots & \vdots & \vdots & & \ddots & \\ 0 & \cdots & 0 & \eta_d & 0 & & \eta_{i+1} \end{bmatrix}, \quad i = 0, 1, \cdots, d-1.$$

$$\underbrace{\phantom{0 \cdots 0}}_{i \text{ columns}} \quad \underbrace{\phantom{2\eta_{i+1} \quad 0}}_{d-i \text{ columns}}$$

Due to space limitations, we provide the technical proof in Appendix A. Lemma 1 shows that by manipulating the nonlinear function vector, $\eta$ can be separated, thereby transferring the nonlinear least squares to a simple linear one. Lemma 1 holds for most of the nonlinear grey models, although only the case of

quadratic nonlinearities is presented here. For higher order polynomials and other forms such as trigono-metric, Lemma 1 provides a reference.

For simplicity of presentation, the transformed parameters at the right-hand side of equations (20) and (21) are denoted as $[\boldsymbol{\vartheta}_\mathrm{L} \; \boldsymbol{\vartheta}_\mathrm{N} \; \boldsymbol{\eta}]$. Sequentially, applying Lemma 1 to equation (19), we can obtain a pseudo-linear regression

$$\boldsymbol{x}(t_k) = \boldsymbol{\vartheta}_\mathrm{L}\tilde{\boldsymbol{x}}(t_k) + \boldsymbol{\vartheta}_\mathrm{N}\boldsymbol{N}\left(\tilde{\boldsymbol{x}}(t_k)\right) + \boldsymbol{\eta} + \boldsymbol{e}(t_k) \tag{22}$$

then, similar to the manipulation in Section 2, one has matrix form

$$\boldsymbol{X} = \Omega(\boldsymbol{x})\Pi + \boldsymbol{E} \tag{23}$$

where

$$\Pi = \begin{bmatrix} \boldsymbol{\vartheta}_\mathrm{L}^\top \\ \boldsymbol{\vartheta}_\mathrm{N}^\top \\ \boldsymbol{\eta}^\top \end{bmatrix}, \; \Omega(\boldsymbol{x}) = \begin{bmatrix} \tilde{\boldsymbol{x}}^\top(t_2) & \boldsymbol{N}^\top\left(\tilde{\boldsymbol{x}}^\top(t_2)\right) & 1 \\ \tilde{\boldsymbol{x}}^\top(t_3) & \boldsymbol{N}^\top\left(\tilde{\boldsymbol{x}}^\top(t_3)\right) & 1 \\ \vdots & \vdots & \vdots \\ \tilde{\boldsymbol{x}}^\top(t_n) & \boldsymbol{N}^\top\left(\tilde{\boldsymbol{x}}^\top(t_n)\right) & 1 \end{bmatrix}, \; \boldsymbol{E} = \begin{bmatrix} \boldsymbol{e}^\top(t_2) \\ \boldsymbol{e}^\top(t_3) \\ \vdots \\ \boldsymbol{e}^\top(t_n) \end{bmatrix}.$$

Minimizing the least-squares objective function $\mathcal{L}(\Pi) = \|\boldsymbol{X} - \Omega(\boldsymbol{x})\Pi\|_\mathrm{F}^2$ gives the simultaneous esti-mates of structural parameters and initial value:

$$\hat{\Pi} = \left(\Omega^\top(\boldsymbol{x})\Omega(\boldsymbol{x})\right)^{-1}\Omega^\top(\boldsymbol{x})\boldsymbol{X}. \tag{24}$$

Then, according to equation (20) or (21), we can obtain the estimates $\left[\hat{\boldsymbol{\theta}}_\mathrm{L} \; \hat{\boldsymbol{\theta}}_\mathrm{N} \; \hat{\boldsymbol{\eta}}\right]$ by inversely solving algebraic equations. Finally, substituting the estimates into the integro-differential equation model (15) gives the time response function

$$\hat{\boldsymbol{x}}(t) = \boldsymbol{f}\left(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\theta}}_\mathrm{L}, \hat{\boldsymbol{\theta}}_\mathrm{N}; t\right)$$

and then by substituting time points $\{t_k\}_{k=1}^{n+r}$, we can obtain the fitting and forecasting values of the original time series $\{\hat{\boldsymbol{x}}(t_k)\}_{k=1}^{n+r}$ directly.

It should be noticed that if $\boldsymbol{N}(\cdot)$ includes a power term, then equations (20) and (21) do not hold any more. In such a case, an alternative strategy to trade off the model accuracy and computation complexity, is using $\boldsymbol{x}(t_1)$ to replace $\boldsymbol{\eta}$ when discretizing $\boldsymbol{N}\left(\boldsymbol{\eta} + \int_{t_1}^{t_k} \boldsymbol{s}(\tau)d\tau\right)$ in equation (17), leading to

$$\boldsymbol{x}(t_k) = \boldsymbol{\theta}_\mathrm{L}\tilde{\boldsymbol{x}}(t_k) + \boldsymbol{\theta}_\mathrm{N}\left[\boldsymbol{N}(\boldsymbol{x}(t_1) + \tilde{\boldsymbol{x}}(t_k)) - \boldsymbol{N}(\boldsymbol{x}(t_1))\right] + \boldsymbol{\eta} + \boldsymbol{e}(t_k)$$

and, by performing the linear least squares criterion, it is straightforward to obtained the structural param-eters and initial value estimates; see [42] for a particular example.

## 4.3. Comparison between nonlinear grey models and integro-differential equation models

In order to analyse the relationship between nonlinear grey system model and its integro-differential equation-based reconstruction, we summarise the modelling procedures of both models in Figure 1.

It can be seen in Figure 1 that the integral operator (or its numerical discretization-based Cusum oper-ator) bridges two models, that is, both use the integral operator (Cusum operator) but in different manner. In particular:
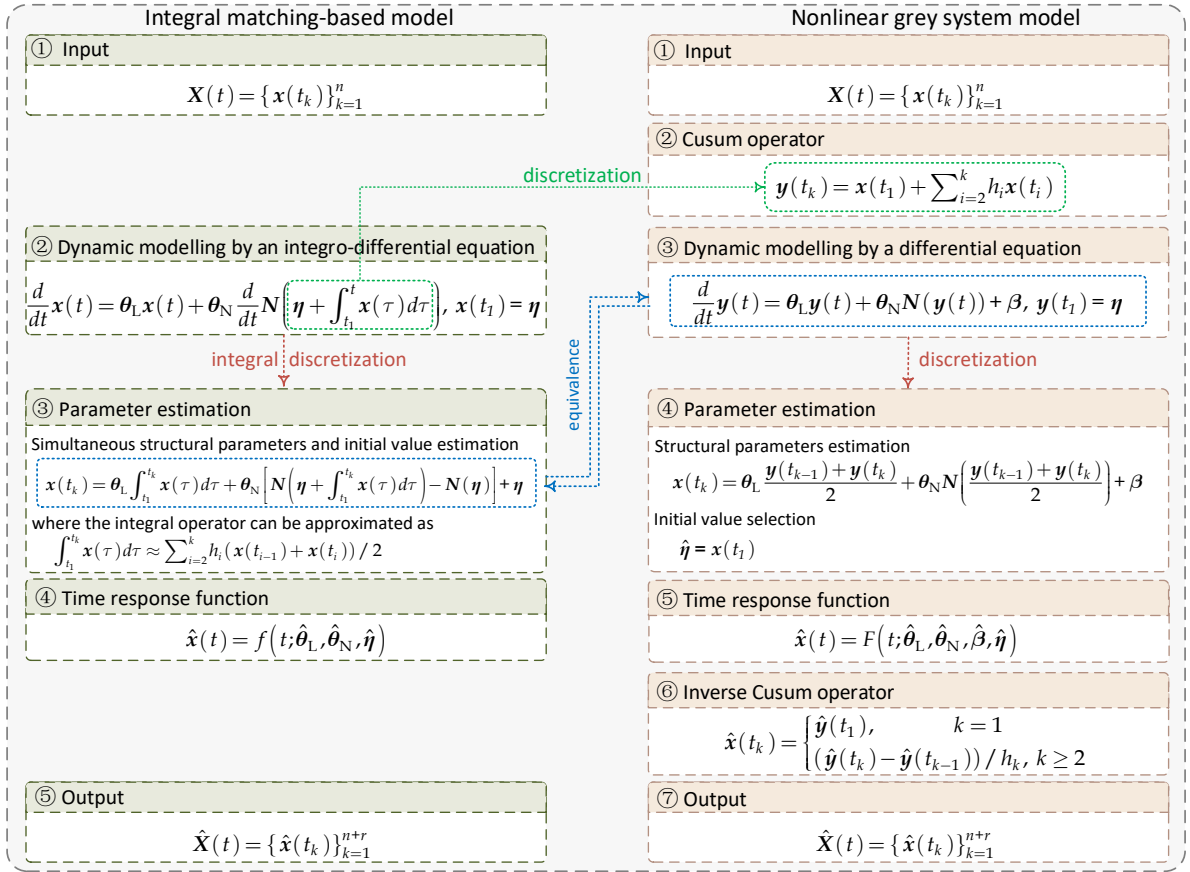
Figure 1: Modelling procedure comparison between nonlinear grey models and integro-differential equation models.

(1) The Cusum operator is the piecewise left-constant quadrature formula of the integral operator. Besides, the existing variations of Cusum operators, such as the fractional cumulative sum operator [40] and new information priority accumulation [39], can be regarded as different numerical discretization forms of the integral operator.

(2) The integro-differential equation can be converted to the nonlinear grey model via a simple integral manipulation, indicating that the nature of the nonlinear grey model is an integro-differential equation model.

(3) It is clear that the modelling process of nonlinear grey models can be viewed as the reordered version of the integral matching-based integro-differential equation models. However, the nonlinear grey modelling requires the inverse cumulative sum step, likely compromising the modelling accuracy.

(4) By applying the integral matching approach to integro-differential equation models, the structural parameters and initial value are obtained simultaneously, whereas the nonlinear grey models need an extra initial value selection strategy (see details in section 2).

## 5. Simulations

In this section we conduct large-scale Monte Carlo experiments to evaluate the finite sample performance of the nonlinear grey system model and it reconstructed one.

Considering the reduced integro-differential equation of the nonlinear grey Verhulst model with $d = 1$, the time series are generated from the state space equation:

$$\text{State equation } \frac{d}{dt}x(t) = ax(t) + bx(t)\left(\eta + \int_{t_1}^{t} x(\tau)d\tau\right),\ x(t_1) = \eta,\ t_1 \geq 0$$

$$\text{Observation equation } x(t_k) = s(t_k) + e(k),\ e(k) \sim \mathcal{N}(0, \sigma^2)$$

where the structural parameters are $a = 1.2$, $b = -1$ and the initial value is $\eta = 0.4$. According to Theorem 1, it is easy to show the corresponding grey Verhulst model is

$$\frac{d}{dt}y(t) = ay(t) + \frac{b}{2}\left(y(t)\right)^2,\ y(t_1) = \eta,\ t \geq 0.$$

### 5.1. Experimental fashions and performance criterion

Since the finite sample performance is affected by the amount of data and measurement noise, they are set to different values. On the one hand, by varying the data scale, we sample data at every time interval of $h$ in the range of $t \in [0, T]$, thereby generating $n = \left\lceil \frac{T}{h} \right\rceil + 1$ samples. On the other hand, we set different magnitudes of Gaussian noise which are controlled by the noise level:

$$\text{Noise Level (\%)} = \frac{\text{var}(\text{Noise})}{\text{var}(\text{Signal})} \times 100\% = \frac{\sigma^2}{\text{var}[s]} \times 100\%,\ i = 1, 2, \cdots, d. \tag{25}$$

Specifically, we consider 2 experimental set-ups. Let $T = 4$. In the first set-up, we set the noise level to 10% but change the time interval $h = [0.40, 0.20, 0.08, 0.04]$, thereby generating $n = [11, 21, 51, 101]$ samples. In the second set-up, we fix the data scale $n = 501$ ($h = 0.01$) but change the noise level $[10\%, 15\%, 20\%, 25\%]$. In each case 500 Monte Carlo realizations are repeated.

For ease of comparison, the parameter and initial value estimation performance are depicted by the violin plot synergistically combining the boxplots and the density trace, where the boxplots shows centre, spread, asymmetry and outliers, and the density trace shows the distributional characteristics of batches of data [11]. In addition, the fitting performance is measured by the mean absolute percentage error criteria

$$\text{RMSE}[x] = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}\left(\hat{x}(t_k) - x(t_k)\right)^2}.$$

In the following, the universal modelling framework of the nonlinear grey system is abbreviated as nonlinear grey modelling, and the reconstructed modelling process is abbreviated as integral matching. Note, the fixed point strategy is employed to select initial conditions for solving the original nonlinear grey model.

### 5.2. Performance evaluation under varying data scales

The estimation of structural parameters and initial values, as well as the fitting errors obtained from integral matching and nonlinear grey modelling are summarised in Figure 2.

Overall, based on the results, we can see that integral matching outperforms nonlinear grey modelling, including parameter estimation accuracy and fitting accuracy. In terms of parameter estimation, with the
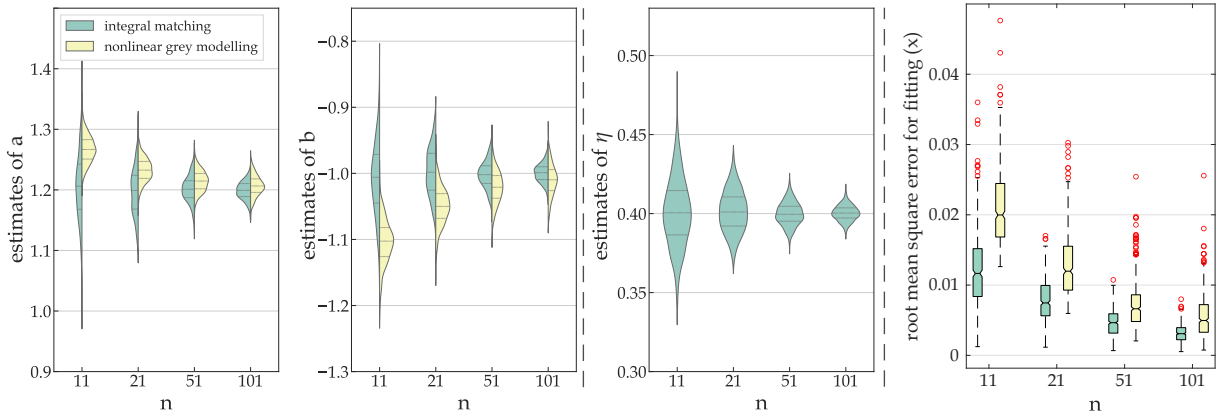
Figure 2: Violin plots of estimated structural parameters, initial value and boxplot of fitting error. The true parameters are $a = 1.2$, $b = -1$ and $\eta = 0.4$, and the nonlinear grey modelling uses $x(t_1)$ as initial value.

increasing of sample size, the estimates of both approaches go to the true $a = 1.2$, $b = -1.0$, suggesting the asymptotic property of integral matching and nonlinear grey modelling. Apparently, $\hat{a}$ and $\hat{b}$ of integral matching tend to have symmetric violin plots, both centring around the true values 1.2 and $-1.0$ in all cases, indicating the likely unbiased estimation of integral matching; whereas the nonlinear grey modelling generates biased estimators even in the case of large data size ($n = 101$) combination. Meanwhile, when encountered samples are small sized, $n = 11$, the variance (measured by the violin shape) of nonlinear grey modelling is smaller than that of integral matching, suggesting the robustness of the former is slightly better than that of the later. As $n$ grows, the variance of both methods decreases, but the decrease of nonlinear grey modelling has a smaller rate than integral matching. As for initial value estimation, the estimates approaches the true value $\eta = 0.4$ as the increase of data scale, validating the efficiency of the proposed estimation method. From the perspective of fitting performance, as the data size increases, the fitting performance of both models ameliorates. Meanwhile, integral matching has lower RMSEs and less outliers in all cases, indicating higher fitting accuracy than nonlinear grey modelling.

## 5.3. Performance evaluation under varying noise levels

The above results demonstrate that the large data scale performs best, thus to probe the effect of noise level, we carry on the experiment in the case of $n = 101$. Figure 3 summarizes the distributions of estimated structural parameters, initial values and fitting errors of integral matching versus nonlinear grey modelling.

As a whole, both integral matching and nonlinear grey modelling perform satisfactory robustness to noise in all noise magnitudes, although the noise impacts the parameter estimation and fitting accuracy. To be specific, from the perspective of parameter estimation accuracy, there only exist slight changes of the medians for both method with the increase of noise level, whereas the distribution of the estimates becomes scattered, indicating the high noise level leads to more outliers. Meanwhile, the two modelling methods depict similar shapes of violin plots, indicating comparable performance of robustness to noise without a substantial difference. Similarly, as for initial value distribution, the increasing noise magnitude brings tiny effect on the median of violin plots for integral matching, while the variance increase obviously. In addition, in the case of large noise level (25%), integral matching still remain higher fitting accuracy, suggesting the efficiency of this kind of modelling method. However, nonlinear grey have some large
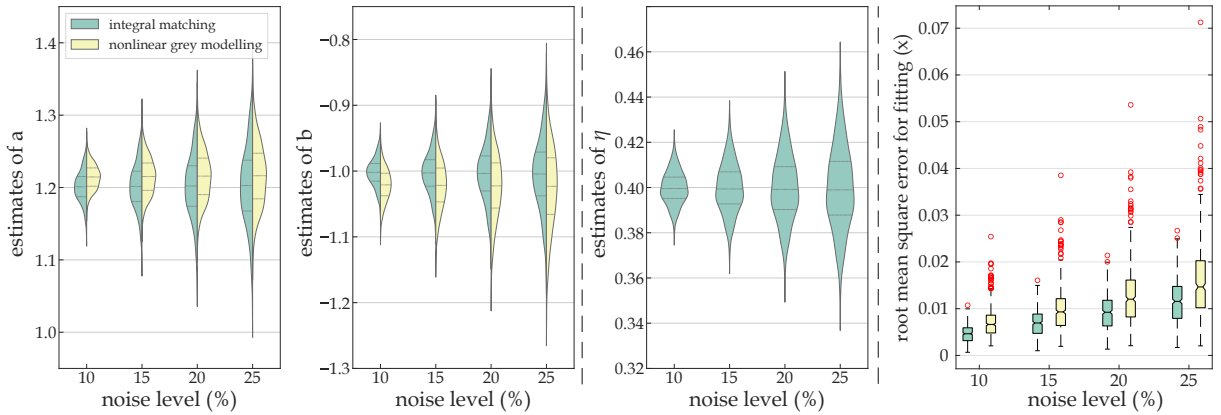
Figure 3: Violin plots of estimated structural parameters, initial value and boxplot of fitting error. The true parameters are $a = 1.2$, $b = -1$ and $\eta = 0.4$, and nonlinear grey modelling uses $x(t_1)$ as initial value.

outliers for the high noise magnitude.

## 5.4. Short discussion

In the single-output scenario, integral matching and nonlinear grey modelling shows several similar results, mainly reflecting in robustness to noise and parameter estimation accuracy for large data size. This is, however, not a general conclusion. In the high-dimensional systems, nonlinear grey modelling performs distinctly worse than integral matching, in terms of parameter estimation accuracy, robustness to noise and fitting accuracy. Take an multi-output system ($d = 2$) as an example, we summarize the modelling results in Appendix B. Particularly, we find that using the first point of the noisy data for solving the original nonlinear grey model leads to blow-up solutions, the reason of which is the loss of accuracy of the numerical method [1], and thus it cannot deal with measurement noise. For illustrating experimental results intuitively, we use the true values of the initial condition instead to solve the 2-dimensional nonlinear grey model.

## 6. Real-world application

Water is vital for human health, industry, agriculture and energy production. Yet with the rapid expansion of urbanization and growth of population, the overall demand for water and the quantity of wastewater produced are both continuously increasing worldwide, which brings formidable threats to the world's water systems [25]. The Yangtze River Delta (YRD) in China, consisting of Shanghai, Jiangsu, Zhejiang, and Anhui Provinces, is one of the key regions for water shortage and pollution due to dense population, fast economic development and high urbanization. Figure 4 shows that the municipal sewage discharge and the total amount water use of the YRD are higher than other regions.

Over the past decades, urban sewage has begun to be comprehensively utilized as a potential water resource, which may alleviate the pressure of the ecological environment and speed the development of the regional economy [33]. Wastewater recovery becomes an efficient measure for water-saving economic patterns. A scientific prediction of municipal sewage is the basis of urban drainage system design, operation and management. Therefore, in this paper, by using the proposed reconstructed nonlinear grey models, we predict the municipal sewage and the total amount of water use to probe the water resources

development of the YRD. The data in the period from 2004 to 2018 is collected from China Statistic Yearbook on Environment.
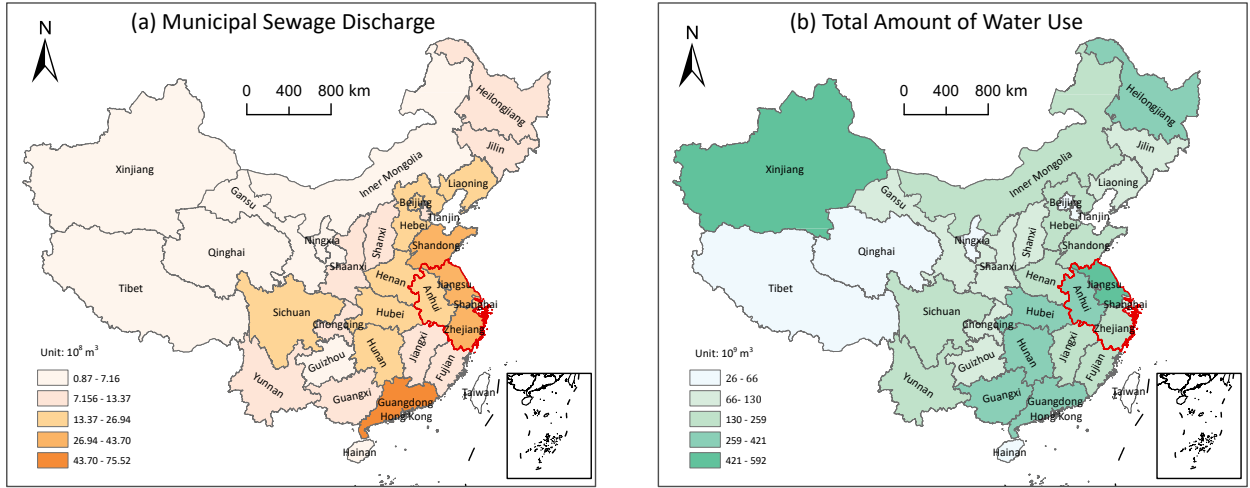


Figure 4: Municipal sewage discharge and total amount of water use of provinces (cities and districts) in 2018.

## 6.1. Results of integro-differential equation models

In order to evaluate the performance of the proposed models, we divide the time series into two parts: 2004-2014 data used to train the models and 2015-2018 data used for assessing the prediction performance. Then, we proceed three-step ahead forecasting for the municipal sewage discharge and water consumption from 2019-2021 to analyse the post-2018 behaviour and support our suggestions.

The models (i)-(iii) in Tables 2 and 3 are three reconstructed integro-differential equation models, almost covering the existing single-output nonlinear grey models (shown as Table 1). And the three models are correspondingly referred to as IGVM, INGM, and INGBM, respectively. INGM and INGBM model have an unknown power term $\gamma$, in the following, we utilize line search [21] to select an appropriate one. Assuming $\gamma \in [a, b]$ on the basis of prior knowledge, there exists $N = \frac{b-a}{\lambda} + 1$ possible choice for $\gamma$, where $\lambda$ is the search step length and, here, set $a = 0$, $b = 2$ and $\lambda = 0.01$. For each possibility, we perform the parameter estimation procedures for the real data and choose the best one (measured by forecasting error). Meanwhile, we use the alternative parameter estimation method for power model families mentioned in subsection 4.2 to address the INGM and INGBM models.

The results obtained from the aforementioned methods are summarized in Tables 2 and 3. From Table 2, we can see that the three integro-differential equation models have similar performance for fitting the municipal sewage discharge data with 2.57%, 2.58% and 2.62% MAPE$_{train}$s, respectively. While the forecasting errors of INGM and INGBM are 0.56% and 0.87%, which is lower than the IGVM (MAPE$_{test}$=2.63%), indicating that models INGM and INGBM are superior to IGVM model. Besides, Table 3 shows that for the total amount of water use scenario, similarly, there are only small differences in the fitting error among the models (i)-(iii). IGVM has the smallest MAPE$_{train}$ for 0.75%, whereas the testing error is the largest one (4.03%) among nonlinear grey models. Apparently, the INGBM shows better forecasting performance than the IGVM and INGM models with MAPE$_{train}$ as low as 1.29%.

To sum up, INGBM performances best for both cases among models (i)-(iii). Therefore, it is probably

the best one to use in this application. In what will follow, we detail the modelling procedures of INGBM model.

Fitting this model on the time series gives the estimated parameters are $a = 0.014$, $b = 0.014$, $\gamma = 1$ and $\eta = 77.33$. The corresponding expression of the INGBM model is $\frac{d}{dt}x(t) = 0.028x(t)$ with $x(t_1) = 77.33$.

Similarly, for the case of total amount of water use, the estimates are $a = -0.06$, $b = 2.89$, $\gamma = 0.63$, $\eta = 1001.65$, and model expression is $\frac{d}{dt}x(t) = -0.06x(t) + 2.89x(t)\left(1001.65 + \int_{t_1}^{t} x(t)\right)^{-0.37}$ with $x(t_1) = 1001.65$.

Table 2: Training and testing results for municipal sewage discharge obztained from reconstructed nonlinear grey models and comparison methods.

| Year | True values ($10^8$ m$^3$) | IGVM[i] | | INGM[ii] | | INGBM[iii] | | ARIMA[iv] | | SVR[v] | | NNAR[vi] | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Values | APE | Values | APE | Values | APE | Values | APE | Values | APE | Values | APE |
| 2004 | 83.00 | 78.85 | 5.00 | 77.92 | 6.12 | 77.33 | 6.84 | – | – | – | – | – | – |
| 2005 | 85.58 | 80.46 | 5.98 | 79.83 | 6.72 | 79.54 | 7.06 | 84.93 | 0.76 | – | – | – | – |
| 2006 | 77.89 | 82.22 | 5.56 | 81.90 | 5.14 | 81.81 | 5.03 | 87.50 | 12.34 | 90.26 | 15.89 | 87.28 | 12.06 |
| 2007 | 80.45 | 84.13 | 4.59 | 84.09 | 4.53 | 84.15 | 4.60 | 79.81 | 0.79 | 80.22 | 0.28 | 80.86 | 0.51 |
| 2008 | 87.27 | 86.23 | 1.19 | 86.40 | 0.99 | 86.55 | 0.82 | 82.37 | 5.61 | 87.05 | 0.24 | 82.07 | 5.96 |
| 2009 | 88.11 | 88.51 | 0.45 | 88.83 | 0.82 | 89.03 | 1.04 | 89.19 | 1.22 | 88.40 | 0.32 | 88.42 | 0.35 |
| 2010 | 92.53 | 90.99 | 1.66 | 91.38 | 1.25 | 91.57 | 1.04 | 90.04 | 2.70 | 92.35 | 0.20 | 90.09 | 2.64 |
| 2011 | 95.08 | 93.70 | 1.44 | 94.04 | 1.09 | 94.19 | 0.93 | 94.46 | 0.65 | 94.30 | 0.82 | 94.10 | 1.03 |
| 2012 | 97.88 | 96.65 | 1.25 | 96.81 | 1.09 | 96.88 | 1.02 | 97.00 | 0.89 | 97.33 | 0.55 | 96.93 | 0.96 |
| 2013 | 99.82 | 99.87 | 0.05 | 99.71 | 0.11 | 99.65 | 0.17 | 99.80 | 0.02 | 99.60 | 0.22 | 99.74 | 0.08 |
| 2014 | 102.24 | 103.38 | 1.11 | 102.72 | 0.47 | 102.50 | 0.25 | 101.74 | 0.49 | 101.62 | 0.60 | 101.83 | 0.40 |
| MAPE$_{train}$ (%) | | | 2.57 | | 2.58 | | 2.62 | | 2.55 | | 2.12 | | 2.67 |
| 2015 | 106.27 | 107.21 | 0.88 | 105.86 | 0.39 | 105.43 | 0.80 | 104.17 | 1.98 | 103.731 | 2.39 | 104.23 | 1.92 |
| 2016 | 110.02 | 111.40 | 1.25 | 109.13 | 0.81 | 108.44 | 1.44 | 106.09 | 3.57 | 105.41 | 4.19 | 106.31 | 3.37 |
| 2017 | 111.40 | 115.98 | 4.11 | 112.53 | 1.01 | 111.54 | 0.12 | 108.01 | 3.04 | 107.06 | 3.90 | 108.41 | 2.69 |
| 2018 | 116.05 | 121.00 | 4.27 | 116.07 | 0.02 | 114.73 | 1.14 | 109.94 | 5.26 | 108.76 | 6.28 | 110.54 | 4.75 |
| MAPE$_{test}$ (%) | | | 2.63 | | 0.56 | | 0.87 | | 3.47 | | 4.79 | | 3.60 |

Note that the model structures are (i) $\frac{d}{dt}x(t) = ax(t) + bx(t)y(t)$; (ii) $\frac{d}{dt}x(t) = bx(t)y^{\gamma-1}(t)$; (iii) $\frac{d}{dt}x(t) = ax(t) + bx(t)y^{\gamma-1}(t)$; (iv) the model order is arima(0,1,0); (v) the embedding dimension is 2 and the kernel type is polynomial; (vi) the model type is feed-forward neural network with 2 lagged inputs, 3 neurons in the only hidden layer.

Table 3: Training and testing results for total amount of water use obtained reconstructed nonlinear grey models and comparison methods.

| Year | True values ($10^9$ m$^3$) | IGVM[i] | | INGM[ii] | | INGBM[iii] | | ARIMA[iv] | | SVR[v] | | NNAR[vi] | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Values | APE | Values | APE | Values | APE | Values | APE | Values | APE | Values | APE |
| 2004 | 1061.25 | 1038.08 | 2.18 | 1040.04 | 2.00 | 1001.65 | 5.62 | – | – | – | – | – | – |
| 2005 | 1058.94 | 1071.36 | 1.17 | 1082.30 | 2.21 | 1064.60 | 0.53 | 1073.66 | 1.17 | – | – | – | – |
| 2006 | 1115.09 | 1101.01 | 1.26 | 1107.61 | 0.67 | 1106.32 | 0.79 | 1012.89 | 4.35 | 1114.31 | 0.07 | 1079.13 | 3.23 |
| 2007 | 1121.56 | 1126.53 | 0.44 | 1125.75 | 0.37 | 1135.37 | 1.23 | 1118.72 | 0.33 | 1134.27 | 1.13 | 1107.78 | 1.23 |
| 2008 | 1161.07 | 1147.49 | 1.17 | 1139.90 | 1.82 | 1155.52 | 0.48 | 1092.05 | 2.62 | 1145.29 | 1.36 | 1135.32 | 2.22 |
| 2009 | 1164.05 | 1163.51 | 0.05 | 1151.51 | 1.08 | 1168.91 | 0.42 | 1168.19 | 0.61 | 1157.22 | 0.59 | 1158.96 | 0.44 |
| 2010 | 1174.64 | 1174.31 | 0.03 | 1161.36 | 1.13 | 1176.93 | 0.20 | 1163.56 | 0.04 | 1178.68 | 0.34 | 1177.50 | 0.24 |
| 2011 | 1173.80 | 1179.69 | 0.50 | 1169.92 | 0.33 | 1180.56 | 0.58 | 1185.58 | 0.93 | 1178.47 | 0.40 | 1184.38 | 0.90 |
| 2012 | 1155.60 | 1179.54 | 2.07 | 1177.47 | 1.89 | 1180.54 | 2.16 | 1202.10 | 2.41 | 1190.61 | 3.03 | 1188.48 | 2.85 |
| 2013 | 1194.20 | 1173.88 | 1.70 | 1184.25 | 0.83 | 1177.45 | 1.40 | 1127.10 | 2.47 | 1204.65 | 0.88 | 1192.83 | 0.11 |
| 2014 | 1162.20 | 1162.79 | 0.05 | 1190.38 | 2.43 | 1171.73 | 0.82 | 1236.30 | 3.52 | 1180.00 | 1.53 | 1191.07 | 2.48 |
| MAPE$_{train}$ (%) | | | 0.75 | | 1.24 | | 0.91 | | 1.62 | | 1.04 | | 1.52 |
| 2015 | 1153.10 | 1146.50 | 0.57 | 1195.99 | 3.72 | 1163.79 | 0.93 | 1172.30 | 0.87 | 1239.50 | 7.49 | 1237.28 | 7.30 |
| 2016 | 1154.00 | 1125.29 | 2.49 | 1201.16 | 4.09 | 1153.93 | 0.01 | 1182.39 | 2.54 | 1238.74 | 7.34 | 1216.65 | 5.43 |
| 2017 | 1165.90 | 1099.54 | 5.69 | 1205.94 | 3.43 | 1142.42 | 2.01 | 1192.49 | 3.34 | 1241.89 | 6.52 | 1259.58 | 8.03 |
| 2018 | 1155.00 | 1069.69 | 7.39 | 1210.40 | 4.80 | 1129.51 | 2.21 | 1202.58 | 3.15 | 1239.07 | 7.28 | 1251.82 | 8.38 |
| MAPE$_{test}$ (%) | | | 4.03 | | 4.01 | | 1.29 | | 2.47 | | 7.05 | | 7.29 |

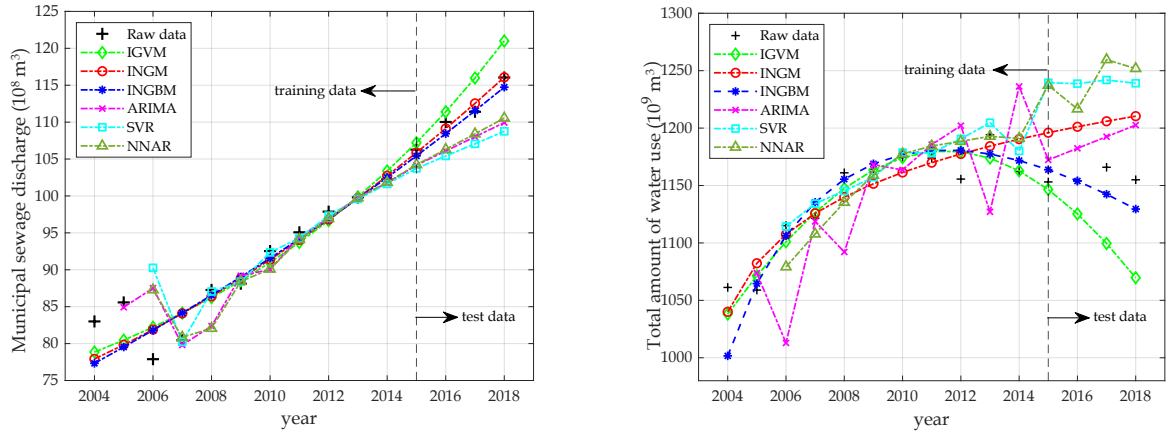Note that all the models in this table share same structures with those in Table 2.

Figure 5: Training and testing results for the YRD yearly municipal sewage discharge of and total amount of water use.

## 6.2. Comparison with other methods

In addition, the reconstructed nonlinear grey models are compared with time series analysis methods, including autoregressive integrated moving average (ARIMA), support vector regression (SVR), and neural network autoregression (NNAR). SVR and NNAR models are implemented by svm and neural_network functions in sklearn package [22], and AIRMA by the arima_model function in statsmodels package [23] in Python software.

Tables 2-3 and Figure 5 depict the comparison between integro-differential equation models and aforementioned methods. The results show that, for both cases, the three comparison methods has satisfactory performance on fitting data with 2.55%, 2.12% and 2.63% $MAPE_{train}$s in the first case (1.62%, 1.04% and 0.77% $MAPE_{train}$s in the second case), which is almost the same as the proposed nonlinear grey models. But their $MAPE_{test}$s vary a lot. Obviously, in the first case, SVR has the best performance in fitting data with $MAPE_{train}$=2.12%, whereas performs poorly in forecasting with $MAPE_{test}$=4.79%. In the case of water consumption, the $MAPE_{train}$ and $MAPE_{test}$ of SVR are 1.04% and 7.05% respectively, indicating high fitting accuracy and worse forecasting accuracy. With regard to NNAR, this result occurs again, which may be over-fitting due to the small data scale. ARIMA considers the autocorrelation, thereby having satisfactory fitting and forecasting performance, although not the optimal one. Overall, the INGBM model can be considered as the most appropriate one in the water resources of the YRD cases.

## 6.3. Short discussion

By employing the INGBM model, we calculate three-step ahead forecasting results of municipal sewage discharge and water consumption as {118.01, 121.38, 124.85} and {1115.4, 1100.2, 1084.2}, respectively. The results show that the total amount of water use in the YRD is expected to decline in the next few years, which is in accordance with the national water control policy. However, with the development of industrialization, the municipal sewage discharge will continue to grow from 2019-2021. Therefore, the government should constantly enhance the municipal sewage treatment capacity from these aspects: (i) expand the scale of sewage treatment plants to relieve the pressure of sewage treatment system, and improve the standard of sewage treatment and reuse; (ii) constantly update the sewage treatment equipment and sewage treatment technology to increase the efficiency of sewage treatment; (iii) further improve the su-

pervision and management, pay attention to improve the management ability and strengthen the real-time supervision ability.

## 7. Conclusions

In this paper, we revisit nonlinear grey system models with an integro-differential equation, including unification, reconstruction, and application. We propose a unified nonlinear grey system model, which covers but not limited to the existing single-variable, multi-variable variable and multi-output models, making it easier for researchers to perform property analysis. It has shown that the unified framework can be reconstructed to an equivalent integro-differential equation, and an integral operator bridges the two models. It has shown that the structural parameters and initial values can be estimated simultaneously by the integral matching approach. The unified framework and a reduced-order integro-differential equation-based model both concern the dynamic modelling, where the former model the Cusum series implicit, whereas the later provide a direct modelling paradigm. The large-scale simulations demonstrate that the reconstructed model is superior to the original ones, from estimation accuracy to robustness to noise. In the practical application, data sets of the municipal sewage discharge and the total amount of water use of the YRD serve to validate the reconstructed models in forecasting for real data.

The present work opens up a new way for the development of nonlinear grey system models, and there are several interesting directions expanding this work: The unified representation (3) is used on the assumption with model structural is known in advance. Determination of model structural from time series in the presence of measurement error is an interesting topic in the future. The modelling mechanism of nonlinear grey models has been invaginated from the perspective of mathematics, however, the superiority of which needs to be further probe in physical analysis and practical applications.

## Appendix

## A. Proof for Lemma 1

*Proof.* Our first destination is to show the case of $d = 1$. Denoting the parameter vector as $\left[\eta \mid \theta_{\mathrm{L}} \mid \theta_{\mathrm{N}}\right] = \left[\eta \mid a \mid b_1 \ \cdots \ b_p\right]$, the left-hand side of equation (20) can be expressed as a algebraic form

$$\eta + a\tilde{x}(t_k) + \sum_{m=1}^{p+1} b_m \left[(\eta + \tilde{x}(t_k))^m - \eta^m\right] \tag{26}$$

then, according to the binomial theorem, the polynomial term can be expanded into

$$(\eta + \tilde{x}(t_k))^m = \sum_{i=1}^{m} \binom{m}{i} \eta^{m-i} \tilde{x}(t_k)^i$$

and thus the problem (26) can be manipulated to

$$\eta + a\tilde{x}(t_k) + b_1 \sum_{i=1}^{2} \binom{2}{i} \eta^{2-i} \tilde{x}(t_k)^i + \cdots + b_p \sum_{i=1}^{p+1} \binom{p+1}{i} \eta^{p+1-i} \tilde{x}(t_k)^i$$

which can be represented as a matrix form

$$
\begin{bmatrix} \eta & a & b_1 & \cdots & b_p \end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 \\
0 & \binom{2}{1}\eta^1 & \binom{2}{2}\eta^0 & & \\
\vdots & \vdots & \vdots & \ddots & \\
0 & \binom{p+1}{1}\eta^p & \binom{p+1}{2}\eta^{q-2} & \cdots & \binom{p+1}{p+1}\eta^0
\end{bmatrix}
\begin{bmatrix}
1 \\
\tilde{x}(t_k) \\
\tilde{x}(t_k)^2 \\
\vdots \\
\tilde{x}(t_k)^q
\end{bmatrix}.
$$

Denoting $\boldsymbol{\phi} = \begin{bmatrix} \binom{2}{1}\eta^1 \\ \vdots \\ \binom{p+1}{1}\eta^p \end{bmatrix}$, and $\boldsymbol{\varphi} = \begin{bmatrix} \binom{2}{2}\eta^0 & & \\ \vdots & \ddots & \\ \binom{p+1}{2}\eta^{q-2} & \cdots & \binom{p+1}{p+1}\eta^0 \end{bmatrix}$ gives the equation (20).

Second, when $d \geq 2$, since $N(\cdot)$ consists of quadratic nonlinearities, it follows that $p = \frac{d(d+1)}{2}$ and, each component of $N(\cdot)$ shares a same form

$$(\eta_i + x_i)(\eta_j + x_j) - \eta_i\eta_j = \eta_i x_j + \eta_j x_i + x_i x_j, \ i, j = 1, 2, \cdots, d.$$

Accordingly, the second term at the right-hand side in equation (19) is transformed into

$$N(\boldsymbol{\eta} + \tilde{x}(t_k)) - N(\boldsymbol{\eta}) = \boldsymbol{\psi}\tilde{x}(t_k) + N(\tilde{x}(t_k)) \tag{27}$$

where $\boldsymbol{\psi} = \begin{bmatrix} \psi_d^\top & \psi_{d-1}^\top & \cdots & \psi_1^\top \end{bmatrix}^\top$, for

$$
\psi_{d-i} = \begin{bmatrix}
0 & \cdots & 0 & 2\eta_{i+1} & & & 0 \\
0 & \cdots & 0 & \eta_{i+2} & \eta_{i+1} & & \\
\vdots & \ddots & \vdots & \vdots & & \ddots & \\
0 & \cdots & 0 & \eta_d & 0 & & \eta_{i+1}
\end{bmatrix}, \ i = 0, 1, \cdots, d-1
$$

$$\underbrace{\phantom{0 \cdots 0}}_{i \text{ columns}} \quad \underbrace{\phantom{\eta_{i+2} \eta_{i+1}}}_{d-i \text{ columns}}$$

Substituting equation (27) into problem (19) gives equation (21). $\square$

## B. Modelling results for a multi-output model

This simulation example is designed to further evaluate the proposed method under a high-dimensional setting. We adopt the reduced nonlinear grey Lotka-Volterra model:

$$
\begin{cases}
\frac{d}{dt}x_1 = a_1 x_1 - b_1 \left[ x_1 \left( \eta_2 + \int_{t_1}^t x_2(\tau)d\tau \right) + x_2 \left( \eta_1 + \int_{t_1}^t x_1(\tau)d\tau \right) \right] \\
\frac{d}{dt}x_2 = a_2 x_2 - b_2 \left[ x_1 \left( \eta_2 + \int_{t_1}^t x_2(\tau)d\tau \right) + x_2 \left( \eta_1 + \int_{t_1}^t x_1(\tau)d\tau \right) \right]
\end{cases}, \quad
\begin{array}{l} x_1(t_1) = \eta_1 \\ x_2(t_2) = \eta_2 \end{array}
$$

Correspondingly, nonlinear grey Lotka-Volterra model is given by

$$
\begin{cases}
\frac{d}{dt}y_1 = a_1 y_1 - b_1 y_1 y_2 \\
\frac{d}{dt}y_2 = a_2 y_2 - b_2 y_1 y_2
\end{cases}, \quad
\begin{array}{l} y_1(t_1) = \eta_1 \\ y_2(t_1) = \eta_2 \end{array}
$$

Here, we set the structural parameters to $a_1 = 1.2$, $b_1 = 0.3$, $a_2 = -1.0$ and $b_2 = -0.4$, the initial value to $\eta_1 = 5.0$ and $\eta_2 = \frac{2}{3}$.

Experiments are conducted in the following fashion. Initially, let $T = 5$ and $h = 0.01$, a number of $n = 501$ samples are generated, following that the noise level $[4\%, 8\%, 12\%, 16\%]$ are considered. In

turn, later, we conduct the experiment with a fixed noise magnitude 4% and varying sample size $n =$ [21, 51, 101, 501]. Simulation results are summarised Figures 6 and 7.
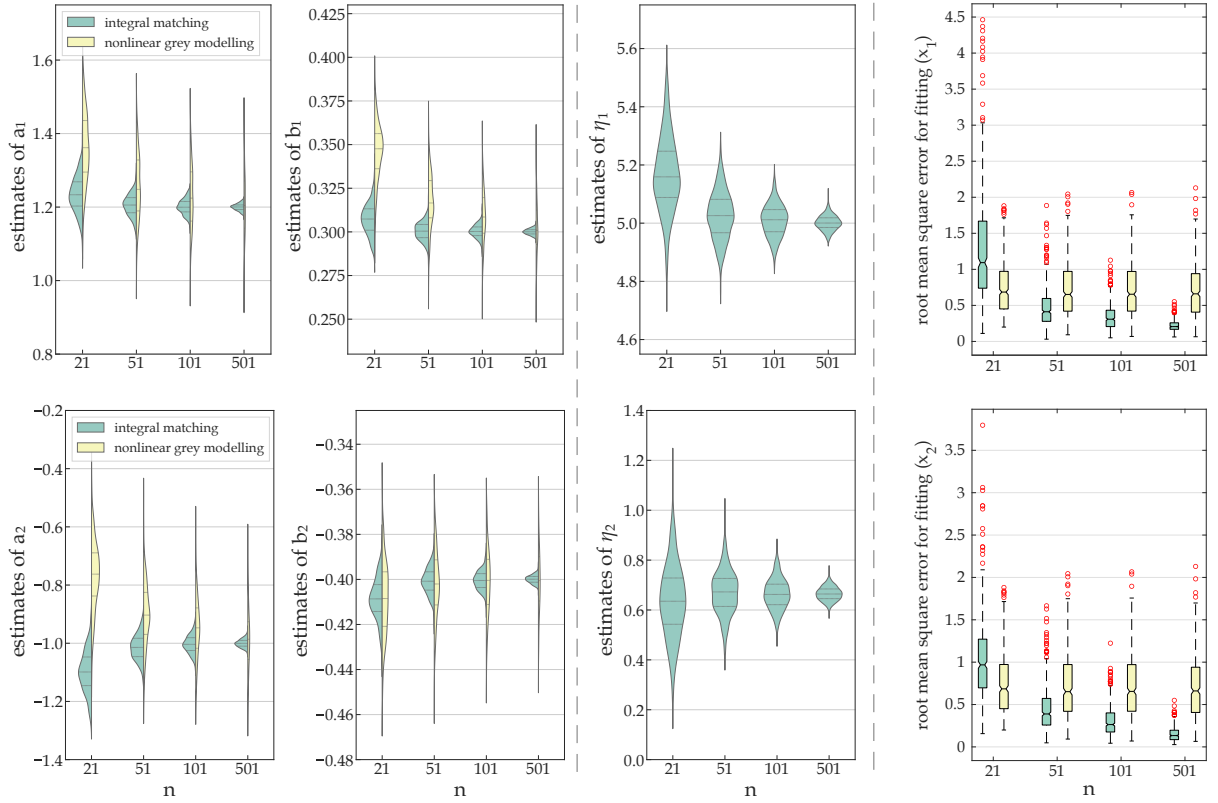


Figure 6: Violin plots of estimated structural parameters, initial value and boxplot of fitting error. The true parameters and initial value are $a_1 = 1.2$, $b_1 = 0.3$, $a_2 = -1$, $b_2 = -0.4$ and $\eta_1 = 5$, $\eta_2 = 2/3$. Note, nonlinear grey modelling uses the true values of the initial condition $\eta_1 = 5$ and $\eta_2 = 2/3$ for solving the time response function.
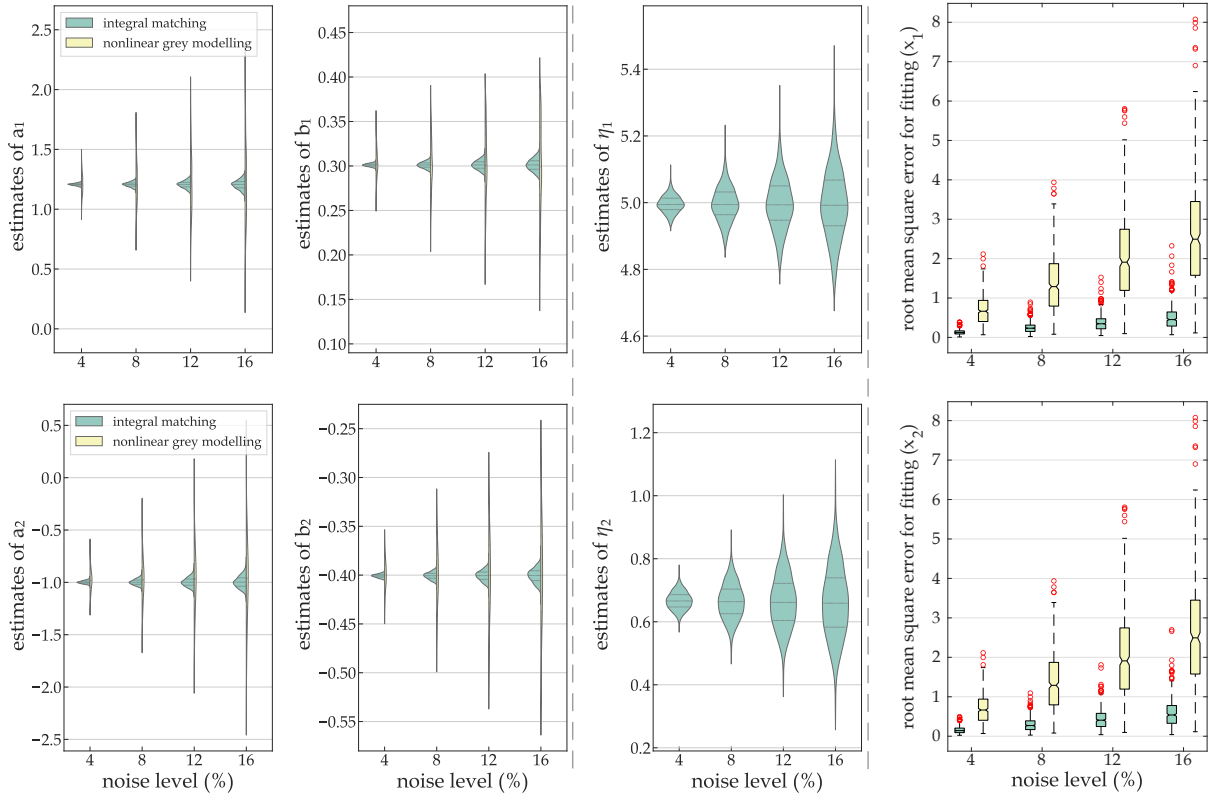
Figure 7: Violin plots of estimated structural parameters, initial value and boxplot of fitting error. The true parameters and initial value are $a_1 = 1.2$, $b_1 = 0.3$, $a_2 = -1$, $b_2 = -0.4$, and $\eta_1 = 5$, $\eta_2 = 2/3$. Note, nonlinear grey modelling uses the true values of the initial condition $\eta_1 = 5$ and $\eta_2 = 2/3$ for solving the time response function.

## Acknowledgment

## Data accessibility

Data and codes are available at `https://github.com/Yanglu0319/Integral-NGM`.

## CRediT authorship contribution statement

**Lu Yang:** Conceptualization, Methodology, Software, Writing - Original draft. **Naiming Xie:** Funding acquisition, Supervision, Writing - review & editing. **Baolei Wei:** Methodology, Validation, Writing - review & editing. **Xiaolei Wang:** Visualization, Writing - review & editing.

## References

[1] Ascher, U.M., Petzold, L.R., 1998. Computer methods for ordinary differential equations and differential-algebraic equations. volume 61. Siam.

[2] Chen, C.I., Chen, H.L., Chen, S.P., 2008. 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, 1194–1204.

[3] Chen, L., Zhang, L.s., 2010. Combination grey verhulst model based on initial value modification. Mathematics in Practice and Theory 11.

[4] Chia-Nan, W., Van-Thanh, P., 2015. An Improved Nonlinear Grey Bernoulli Model Combined with Fourier Series. Mathematical Problems in Engineering 2015, 1–7.

[5] Dattner, I., Miller, E., Petrenko, M., Kadouri, D.E., Jurkevitch, E., Huppert, A., 2017. Modelling and parameter inference of predator–prey dynamics in heterogeneous environments using the direct integral approach. Journal of The Royal Society Interface 14, 20160525.

[6] Deng, J., 1984. Grey theory and methods in social and economic system. Social Sciences in China 6, 47–60.

[7] Ding, S., Li, R., Wu, S., 2021. A novel composite forecasting framework by adaptive data preprocessing and optimized nonlinear grey bernoulli model for new energy vehicles sales. Communications in Nonlinear Science and Numerical Simulation 99, 105847.

[8] Duan, H., Luo, X., 2020. Grey optimization verhulst model and its application in forecasting coal-related co 2 emissions. Environmental Science and Pollution Research 27, 43884–43905.

[9] Duman, G.M., Kongar, E., Gupta, S.M., 2019. Estimation of electronic waste using optimized multivariate grey models. Waste Management 95, 241–249.

[10] Evans, M., 2014. 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, 1236–1244.

[11] Fisher, N.I., Switzer, P., 2001. Statistical computing and graphics. American Statistician 56, 121–130.

[12] Gao, M., Yang, H., Xiao, Q., Goh, M., 2021. A novel fractional grey Riccati model for carbon emission prediction. Journal of Cleaner Production 282, 124471.

[13] Gatabazi, P., Mba, J., Pindza, E., Labuschagne, C., 2019a. Grey Lotka–Volterra models with application to cryptocurrencies adoption. Chaos, Solitons & Fractals 122, 47–57.

[14] Gatabazi, P., Mba, J.C., Pindza, E., 2019b. Fractional gray Lotka-Volterra models with application to cryptocurrencies adoption. Chaos: An Interdisciplinary Journal of Nonlinear Science 29, 073116.

[15] Hsu, L.C., 2010. A genetic algorithm based nonlinear grey Bernoulli model for output forecasting in integrated circuit industry. Expert Systems with Applications 37, 4318–4323.

[16] Hyndman, R.J., Koehler, A.B., Snyder, R.D., Grose, S., 2002. A state space framework for automatic forecasting using exponential smoothing methods. International Journal of forecasting 18, 439–454.

[17] King, R., 2012. A review of bayesian state-space modelling of capture–recapture–recovery data. Interface Focus 2, 190–204.

[18] Ma, X., Liu, Z.b., 2018. The kernel-based nonlinear multivariate grey model. Applied Mathematical Modelling 56, 217–238.

[19] Mao, S., Gao, M., Zhu, M., 2015. The impact of R&D on GDP study based on grey delay Lotka-Volterra model. Grey Systems: Theory and Application 5, 74–88.

[20] Mattheij, R., Molenaar, J., 2002. Ordinary differential equations in theory and practice. SIAM.

[21] Nocedal, J., Wright, S., 2006. Numerical optimization. Springer Science & Business Media.

[22] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in python. the Journal of machine Learning research 12, 2825–2830.

[23] Seabold, S., Perktold, J., 2010. Statsmodels: Econometric and statistical modeling with python, in: Proceedings of the 9th Python in Science Conference, Austin, TX. p. 61.

[24] Shaikh, F., Ji, Q., Shaikh, P.H., Mirjat, N.H., Uqaili, M.A., 2017. Forecasting China's natural gas demand based on optimised nonlinear grey models. Energy 140, 941–951.

[25] Srinivasan, V., Lambin, E.F., Gorelick, S.M., Thompson, B.H., Rozelle, S., 2012. The nature and causes of the global water crisis: Syndromes from a meta-analysis of coupled human-water studies. Water Resources Research 48.

[26] Stoica, P., Jansson, M., 2000. Mimo system identification: State-space and subspace approximations versus transfer function and instrumental variables. IEEE Transactions on Signal Processing 48, 3087–3099.

[27] Tang, L., Lu, Y., 2020. Study of the grey Verhulst model based on the weighted least square method. Physica A: Statistical Mechanics and its Applications 545, 123615.

[28] Wang, Q., Li, S., Li, R., Ma, M., 2018. Forecasting us shale gas monthly production using a hybrid arima and metabolic nonlinear grey model. Energy 160, 378–387.

[29] Wang, Q., Song, X., 2019. Forecasting China's oil consumption: A comparison of novel nonlinear-dynamic grey model (GM), linear GM, nonlinear GM and metabolism GM. Energy 183, 160–171.

[30] Wang, Y., Yang, S., Qian, W., Li, X., 2013. Forecasting new product diffusion using grey time-delayed verhulst model. Journal of Applied Mathematics 2013.

[31] Wang, Z., Li, Q., 2019. Modelling the nonlinear relationship between co2 emissions and economic growth using a pso algorithm-based grey verhulst model. Journal of Cleaner Production 207, 214–224.

[32] Wang, Z.X., Zhu, H.T., 2016. Testing the trade relationships between China, Singapore, Malaysia and Thailand using grey Lotka-Volterra competition model. Kybernetes 45, 931–945.

[33] Water, U., 2017. 2017 un world water development report: Wastewater the untapped resource.

[34] Wei, B., Xie, N., 2020a. On unified framework for continuous-time grey models: an integral matching perspective. arXiv preprint arXiv:2008.00522 .

[35] Wei, B., Xie, N., 2020b. On unified framework for discrete-time grey models: Extensions and applications. ISA transactions 107, 1–11.

[36] Wei, B., Xie, N., Yang, L., 2020. Understanding cumulative sum operator in grey prediction model with integral matching. Communications in Nonlinear Science and Numerical Simulation 82, 105076.

[37] Wei, B.l., Xie, N.m., Yang, Y.j., 2019. Data-based structure selection for unified discrete grey prediction model. Expert Systems

with Applications 136, 264–275.

[38] Wu, L., Liu, S., Wang, Y., 2012. Grey Lotka–Volterra model and its application. Technological Forecasting and Social Change 79, 1720–1730.

[39] Wu, L., Zhang, Z., 2018. Grey multivariable convolution model with new information priority accumulation. Applied Mathematical Modelling 62, 595–604.

[40] Wu, W., Ma, X., Zeng, B., Wang, Y., Cai, W., 2019. Forecasting short-term renewable energy consumption of China using a novel fractional nonlinear grey Bernoulli model. Renewable Energy 140, 70–87.

[41] Xiao, Q., Gao, M., Xiao, X., Goh, M., 2020. A novel grey Riccati–Bernoulli model and its application for the clean energy consumption prediction. Engineering Applications of Artificial Intelligence 95, 103863.

[42] Yang, L., Xie, N., 2021. Integral matching-based nonlinear grey Bernoulli model for forecasting the coal consumption in China. Soft Computing 25, 5209–5223.

[43] Young, P.C., 2011. Gauss, kalman and advances in recursive parameter estimation. Journal of Forecasting 30, 104–146.

[44] Yu, L., Ma, X., Wu, W., Wang, Y., Zeng, B., 2021. A novel Elastic Net-based NGBMC(1,n) model with multi-objective optimization for nonlinear time series forecasting. Communications in Nonlinear Science and Numerical Simulation , 105696.

[45] Yuan, Y., 2020. A SAFSA- and Metabolism-Based Nonlinear Grey Bernoulli Model for Annual Water Consumption Prediction. Iranian Journal of Science and Technology , 11.

[46] Zhou, W., Pei, L., 2020. The grey generalized Verhulst model and its application for forecasting Chinese pig price index. Soft Computing 24, 4977–4990.