Particle filter-based algorithm of simultaneous output and parameter estimation for output nonlinear systems under low measurement rate constraints

Mengting Chen · Rongming Lin ·
Teng Yong Ng · Feng Ding

Abstract In applications of system identification where the inputs and outputs are scheduled at different sampling rates, the traditional gradient-based iterative (GI) identification scheme can be applied to approximate the unknown outputs from which the unknown model parameters can be estimated. A limitation of the GI method lies in the fact that the measured outputs are not employed effectively to adjust/improve the missing output estimates. To address this, an improved GI method using the particle filters is designed to jointly estimate the outputs and parameters of output nonlinear systems from the dual-rate data. The key idea is to use a bank of weighted particles to represent the posterior probability density function of the unknown outputs. The kernel density estimation method is then developed to update the weights of these particles at each iteration. The negative gradient search principle and the parameter separation technique are subsequently combined to obtain the required parameter estimation of the plant model. A numerical example including comparisons with the existing algorithms is reported to demonstrate the effectiveness and limitations of the proposed methodology.

M. Chen
School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210003, People's Republic of China
e-mail: mtchen11@126.com; mtchen@njupt.edu.cn

F. Ding (✉)
School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, People's Republic of China
e-mail: fding@jiangnan.edu.cn

R. Lin · T. Y. Ng
School of Mechanical and Aerospace Engineering, College of Engineering, Nanyang Technological University, Singapore 639798, Singapore
e-mail: mrmlin@ntu.edu.sg
T.Y. Ng
e-mail: mtyng@ntu.edu.sg

1 Introduction

Advanced process control strategies are often based on system models, and system identification hence accounts for a significant proportion of applications in modern control theory and practice [1–3]. The basic idea is to compare the system responses with the identification model in some optimal ways in order to confirm how close the model response fits the system response [4–6]. In recent years, nonlinear processes have attracted a lot of attention both from the control and the signal processing point of view and have been increasingly deployed [7–9]. When tackling the problem of modeling nonlinear processes from the observed data, one of the main difficulties is how to manage the trade-off between the model flexibility and the risk of overfitting [10,11].

In engineering practice, nonlinear systems with the multi-rate sampling are ubiquitous [12]. The traditional discrete-time systems which have the same sampling rate among all variables are generally referred to as the single-rate sampled-data systems. However, because of the practical difficulties such as hardware restrictions and the instability of communication networks, it is often impossible to achieve the acquisitions of all signals involved in the process at the same sampling rate [13]. On the other hand, in the discretization process of control systems, the intentional multi-rate sampling of different signals can in fact more accurately capture their potential characteristics and predict better the unknown underlying dynamics. In other words, the application of different sampling rates is desirable and even inevitable in practice [14]. As a result, the research on multi-rate sampled-data systems has recently received considerable and increasing attention over the years [15,16]. Li and Zhang designed the maximum likelihood multi-innovation method to identify the Hammerstein system with the dual-rate sampling [17]; Yan *et al*. introduced a mixed-rate model-based least squares estimator for the multi-rate discrete-time closed-loop system [18]; Chen *et al*. analyzed the multi-rate stabilization problem for the singularly perturbed system by using a time-dependent Lyapunov function [19]. In this paper, the dual-rate nonlinear systems with bilinear forms are considered, which can be regarded as a special case of missing output systems. In the context of the dual-rate sampling, the fast inputs and slow outputs are measured at different sampling rates, with sampling intervals being periodic. The objective of this paper is to reconstruct the inter-sample outputs and to estimate the unknown system parameters based on the inferential control idea.

In the field of the discrete-time system identification, many different identification ideas have been proposed in the literature [20–22]. Among them, the iterative method is attractive as it leads to good estimates of the unknown terms, even when a small amount of data is used for the estimation of the statistics [23–29]. Iterative methods are widely used in gradient-based optimization algorithms such as the auxiliary model gradient-based iterative identification method for multi-input nonlinear output-error systems [30,31]. Another popular tool is the particle filter that can be used to identify nonlinear state-space systems with unknown states [32–34]. Based on the Monte Carlo sampling principle, the posterior probability density function of the state is represented by using a set of the weighted particles [35,36]. Inspired by this, Chen *et al*. developed a new particle filter-based approach to identify a dual-rate nonlinear model by treating the unknown outputs as the states [37]. Later in [38], by combining the multi-innovation identification principle, the particle filter was further extended to the estimation of the unknown process outputs and parameters for the nonstate-space output-error type models.

The plant model to be studied in this paper has a particular bilinear form called the bilinear-in-parameter model, which has been found to be increasingly popular in modeling nonlinear processes [39,40]. When the output nonlinearity of the block-oriented systems can be expressed as a linear combination of a set of the known basis functions of the past input/output data, the system can be transformed into the bilinear-in-parameter system. For example, the bilinear-in-parameter system can be used to describe the Hammerstein–Wiener system [41]. In this paper, we seek to propose identification algorithms for combined output and parameter estimation under low measurement rate constraints. In order to obtain the pseudo-linear regression identification models, two different techniques of tackling the special bilinear term are adopted, namely the over-parameterization technique and the parameter separation technique. Based on the two approximate discrete-time models, the corresponding particle filter-based identification methods are developed for bilinear-in-parameter models in the dual-rate framework. Then, the original parameter estimation problem is decomposed into two sub-problems: i) using the particle filters to recover the missing measurements from the available observations and ii) using the negative gradient search principle [42] to perform parameter estimation based on the recovered output signals. To be more specific, the particle filter is introduced that combines the Epanechnikov function to approximately estimate the missing outputs between the measured output samples. On this basis, the modified gradient-based iterative algorithm can then be employed to identify all the model parameters. As compared with the traditional gradient-based iterative (GI) approaches, the former can improve the accuracy of parameter estimation.

The rest of this paper is organized as follows. Section 2 describes the problem formulation of the dual-rate bilinear-in-parameter models and briefly derives two typical GI algorithms. Section 3 derives the improved GI algorithms by combining the particle fil-

ter. In Sect. 4, a representative numerical example is given to illustrate the effectiveness of the proposed algorithms. The paper concludes with the final remarks in Sect. 5.

2 Problem description

The notations used in this paper are fairly standard. The variables in bold represent multi-dimensional vectors or matrices. $\hat{A}_k$ denotes the iterative estimate of A at k-th iteration. The superscript "T" is the vector/matrix transpose. "$A =: X$" or "$X := A$" reads as "A is defined as X." The Frobenius norm of a matrix A is defined by $\|A\|^2 := \text{tr}[AA^{\text{T}}]$, where $\text{tr}[\cdot]$ denotes the trace of AA^{T}. $\mathbf{1}_n$ represents an n-dimensional column vector with all elements unity. In this section, we first present the bilinear-in-parameter models with the available input–output data. Then, two techniques for processing the bilinear parts of the system are given from the perspective of parameter estimation. On this basis, the corresponding GI identification methods are derived for comparison.

2.1 The dual-rate bilinear-in-parameter system

Consider the following discrete-time output nonlinear systems:

$$y(l) = \sum_{j=1}^{n} \beta_j \sum_{i=1}^{m} \alpha_i g_i(y(l-j)) + \sum_{\iota=1}^{n_\gamma} \gamma_\iota u(l-\iota) + v(l),$$
(1)

where l is the discrete-time index, $u(l) \in \mathbb{R}$ is the system input, $y(l) \in \mathbb{R}$ is the system output, $\{\beta_j\}_{j=1}^n$, $\{\alpha_i\}_{i=1}^m$ and $\{\gamma_\iota\}_{\iota=1}^{n_\gamma}$ are the unknown parameters, $v(l) \in \mathbb{R}$ is a stochastic white noise with zero mean and variance σ^2, and $g_i(\cdot)$, $i = 1, 2, \cdots, m$ are the known nonlinear basis functions.

Define the parameter vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, the information vector $\boldsymbol{\varphi}_u(l)$ and the information matrix $G(l)$ as

$\boldsymbol{\alpha} := [\alpha_1, \alpha_2, \cdots, \alpha_m]^{\text{T}} \in \mathbb{R}^m$, $\quad \boldsymbol{\beta} := [\beta_1, \beta_2, \cdots, \beta_n]^{\text{T}} \in \mathbb{R}^n$,
$\boldsymbol{\gamma} := [\gamma_1, \gamma_2, \cdots, \gamma_{n_\gamma}]^{\text{T}} \in \mathbb{R}^{n_\gamma}$,
$\boldsymbol{\varphi}_u(l) := [u(l-1), u(l-2), \cdots, u(l-n_\gamma)]^{\text{T}} \in \mathbb{R}^{n_\gamma}$,

$$G(l) := \begin{bmatrix} g_1(y(l-1)) & g_1(y(l-2)) & \cdots & g_1(y(l-n)) \\ g_2(y(l-1)) & g_2(y(l-2)) & \cdots & g_2(y(l-n)) \\ \vdots & \vdots & & \vdots \\ g_m(y(l-1)) & g_m(y(l-2)) & \cdots & g_m(y(l-n)) \end{bmatrix} \in \mathbb{R}^{m \times n}.$$
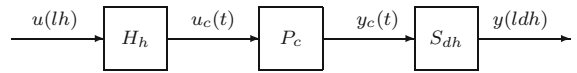

Fig. 1 The dual-rate system

According to the above definitions, the output nonlinear system in (1) can be converted into the following bilinear-in-parameter models:

$$y(l) = \boldsymbol{\alpha}^{\text{T}} G(l) \boldsymbol{\beta} + \boldsymbol{\varphi}_u^{\text{T}}(l) \boldsymbol{\gamma} + v(l).$$
(2)

Remark 1 The bilinear-in-parameter model can describe the output nonlinear system, in which its output nonlinearity is represented as a linear combination of a set of the known basis functions of the past output data. Likewise, if we define $g_i(y(l-j)) := g_i(u(l-j))$, the corresponding input nonlinear system can also readily be converted into the bilinear-in-parameter model [43].

The identification approaches pursued in this paper are based on the assumption of sampling the plant output at a rate slower than the control input, namely, the inputs and outputs are sampled by two different rates, with the former being higher than the latter. Let the measurement (output) sampling period dh be a multiple of h for some integer $d > 1$ and $h > 0$ is the fundamental (input) sampling period. Then the input–output data $\{u(lh), y(ldh)\}$ can be sampled. The detailed configuration is shown in Fig. 1, where H_h is a zero-order hold with period h, S_{dh} a sampler with period dh. The input $u_c(t)$ is generated by H_h, processing a discrete-time signal $u(lh)$; P_c is a continuous-time process. The output of P_c is $y_c(t)$ which is further down-sampled by the sampler S_{dh}, yielding the measured output $y(ldh)$ with period dh. For convenience, letting $h = 1$ and replacing l with ld in (2) result in

$$y(ld) = \boldsymbol{\alpha}^{\text{T}} G(ld) \boldsymbol{\beta} + \boldsymbol{\varphi}_u^{\text{T}}(ld) \boldsymbol{\gamma} + v(ld),$$
(3)

where

$$G(ld) = \begin{bmatrix} g_1(y(ld-1)) & g_1(y(ld-2)) & \cdots & g_1(y(ld-n)) \\ g_2(y(ld-1)) & g_2(y(ld-2)) & \cdots & g_2(y(ld-n)) \\ \vdots & \vdots & & \vdots \\ g_m(y(ld-1)) & g_m(y(ld-2)) & \cdots & g_m(y(ld-n)) \end{bmatrix},$$
(4)

$$\boldsymbol{\varphi}_u(ld) = [u(ld-1), u(ld-2), \cdots, u(ld-n_\gamma)]^{\text{T}}.$$
(5)

The proposed parameter estimation algorithms in this paper are based on this identification model in (3)–(5). Many identification methods are derived based on the identification models of the systems [44–49] and

can be used to estimate the parameters of other linear systems and nonlinear systems [50–55] and can be applied to other fields [56–64] such as chemical process control systems and engineering application systems. Throughout this paper, we refer to (3)–(5) as the fast-rate model. The measured input–output data are $\{u(l), l = 1, 2, \cdots, L\}$ at the fast rate and $\{y(ld), d = 1, 2, \cdots\}$ at the slow rate. Thus, the inter-sample outputs (missing outputs) $\{y(ld - i), i = 1, 2, \cdots, d - 1\}$ are not available. In what follows, we will show how to estimate the unknown parameters by using the GI method with the dual-rate measurement data which are $\{u(l), y(ld)\}$.

2.2 The GI algorithm using the over-parameterization technique

As we can observe, $y_0(ld) := \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{G}(ld) \boldsymbol{\beta}$ is the bilinear function of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. A common approach in statistics for estimation problems of this kind is to over-parameterize the model so that it becomes linear. For simplicity and without loss of generality, we let $\alpha_1 = 1$ for the rest of the paper. This ensures the uniqueness of the identification results. Then, the bilinear part $\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{G}(ld) \boldsymbol{\beta}$ can be written as follows:

$$\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{G}(ld) \boldsymbol{\beta} = \boldsymbol{\varphi}_1^{\mathrm{T}}(ld) \boldsymbol{\rho}, \quad \boldsymbol{\varphi}_1(ld) := \mathrm{col}[\boldsymbol{G}^{\mathrm{T}}(ld)] \in \mathbb{R}^{mn},$$
$$\boldsymbol{\rho} := \boldsymbol{\alpha} \otimes \boldsymbol{\beta}$$
$$= [\alpha_1 \beta_1, \alpha_1 \beta_2, \cdots, \alpha_1 \beta_n, \alpha_2 \beta_1, \alpha_2 \beta_2, \cdots,$$
$$\alpha_2 \beta_n, \cdots, \alpha_m \beta_1, \alpha_m \beta_2, \cdots, \alpha_m \beta_n]^{\mathrm{T}}$$
$$= [\beta_1, \beta_2, \cdots, \beta_n, \alpha_2 \beta_1, \alpha_2 \beta_2, \cdots, \alpha_2 \beta_n,$$
$$\cdots, \alpha_m \beta_1, \alpha_m \beta_2, \cdots, \alpha_m \beta_n]^{\mathrm{T}} \in \mathbb{R}^{mn},$$

where $\otimes$ is the Kronecker product operator and the col$[\cdot]$ operator is one which forms a vector from a matrix by stacking its columns on top of one another.

From the above derivation, the model in (3)–(5) can then be linearized as the following regression-like model:

$$y(ld) = \boldsymbol{\varphi}_1^{\mathrm{T}}(ld) \boldsymbol{\rho} + \boldsymbol{\varphi}_u^{\mathrm{T}}(ld) \boldsymbol{\gamma} + v(ld)$$
$$= \boldsymbol{\phi}^{\mathrm{T}}(ld) \boldsymbol{\theta} + v(ld), \qquad (6)$$

in which

$$\boldsymbol{\phi}(ld) := [\boldsymbol{\varphi}_1^{\mathrm{T}}(ld), \boldsymbol{\varphi}_u^{\mathrm{T}}(ld)]^{\mathrm{T}} \in \mathbb{R}^{mm + n_\gamma},$$
$$\boldsymbol{\theta} := [\boldsymbol{\rho}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{mm + n_\gamma}.$$

Regression-like model (6) cannot be considered as a basis for the direct estimation because it is a pseudo-linear regression equation involving the information vector $\boldsymbol{\phi}(ld)$ that contains the unmeasured inter-sample output data. Thus the typical GI algorithm cannot be used directly to identify the parameter vector $\boldsymbol{\theta}$. Here, we apply the gradient-based iterative algorithm through the over-parameterization technique to estimate $\boldsymbol{\theta}$, which is abbreviated as the O-GI algorithm shown in the following:

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k-1} + \mu_{1,k} \sum_{l=1}^{L} \hat{\boldsymbol{\phi}}_k(ld)[y(ld) - \hat{\boldsymbol{\phi}}_k^{\mathrm{T}}(ld) \hat{\boldsymbol{\theta}}_{k-1}], \quad (7)$$

$$\mu_{1,k} \leq 2\lambda_{\max}^{-1} \left[\sum_{l=1}^{L} \hat{\boldsymbol{\phi}}_k(ld) \hat{\boldsymbol{\phi}}_k^{\mathrm{T}}(ld) \right], \qquad (8)$$

$$\hat{\boldsymbol{\phi}}_k(ld) = [\hat{\boldsymbol{\varphi}}_{1,k}^{\mathrm{T}}(ld), \boldsymbol{\varphi}_u^{\mathrm{T}}(ld)]^{\mathrm{T}}, \qquad (9)$$

$$\hat{\boldsymbol{\varphi}}_{1,k}(ld) = [g_1(\hat{y}_k(ld - 1)), \cdots, g_1(y(ld - d)), \cdots,$$
$$g_1(\hat{y}_k(ld - n)), g_2(\hat{y}_k(ld - 1)), \cdots, g_2(y(ld - d)),$$
$$\cdots, g_2(\hat{y}_k(ld - n)), \cdots, g_m(\hat{y}_k(ld - 1)),$$
$$\cdots, g_m(y(ld - d)), \cdots, g_m(\hat{y}_k(ld - n))]^{\mathrm{T}}, \qquad (10)$$

$$\boldsymbol{\varphi}_u(ld) = [u(ld - 1), u(ld - 2), \cdots, u(ld - n_\gamma)]^{\mathrm{T}}, \quad (11)$$

$$\hat{y}_k(ld - i) = \hat{\boldsymbol{\phi}}_k^{\mathrm{T}}(ld - i) \hat{\boldsymbol{\theta}}_k, \qquad (12)$$

$$\hat{\boldsymbol{\theta}}_k = [\hat{\boldsymbol{\beta}}_k^{\mathrm{T}}, \widehat{\alpha_2 \boldsymbol{\beta}^{\mathrm{T}}}_k, \cdots, \widehat{\alpha_m \boldsymbol{\beta}^{\mathrm{T}}}_k, \hat{\boldsymbol{\gamma}}_k^{\mathrm{T}}]^{\mathrm{T}}$$
$$= [\hat{\beta}_{1,k}, \cdots, \hat{\beta}_{n,k}, \widehat{\alpha_2 \beta_{1,k}}, \cdots, \widehat{\alpha_2 \beta_{n,k}}, \cdots, \widehat{\alpha_m \beta_{1,k}},$$
$$\cdots, \widehat{\alpha_m \beta_{n,k}}, \hat{\boldsymbol{\gamma}}_k^{\mathrm{T}}]^{\mathrm{T}}, \qquad (13)$$

$$\hat{\alpha}_{i,k} = \frac{1}{n} \sum_{j=1}^{n} \frac{\widehat{\alpha_i \beta_{j,k}}}{\hat{\beta}_{j,k}}, \quad i = 2, 3, \cdots, m, \quad j = 1, 2, \cdots, n,$$
$$\qquad (14)$$

$$\hat{\boldsymbol{\vartheta}}_k = [\hat{\boldsymbol{\beta}}_k^{\mathrm{T}}, \hat{\alpha}_{2,k}, \hat{\alpha}_{3,k}, \cdots, \hat{\alpha}_{m,k}, \hat{\boldsymbol{\gamma}}_k^{\mathrm{T}}]^{\mathrm{T}}. \qquad (15)$$

Note that the estimate $\hat{\boldsymbol{\beta}}_k$ can be obtained directly from the first n terms of $\hat{\boldsymbol{\theta}}_k$ since $\alpha_1 = 1$. From (13), we get n values of the parameter estimate $\hat{\alpha}_{i,k}$ and take their average as the value of α_i [40].

2.3 The GI algorithm using the parameter separation technique

Considering the possible problems caused by the product term $a_i b_j$ in the above O-GI algorithm, a parameter separation technique is developed here to deal with the bilinear part. Let $\boldsymbol{g}_i(ld) \in \mathbb{R}^{1 \times n}$ be the i-th row of the information matrix $\boldsymbol{G}(ld)$. The details are listed as follows:

$$\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{G}(ld) \boldsymbol{\beta} = g_1(y(ld - 1)) \beta_1 + \alpha_2 g_2(y(ld - 1)) \beta_1$$
$$+ \cdots + \alpha_m g_m(y(ld - 1)) \beta_1$$

$$+g_1(y(ld-2))\beta_2 + \alpha_2 g_2(y(ld-2))\beta_2$$
$$+\cdots+\alpha_m g_m(y(ld-2))\beta_2 + \cdots$$
$$+g_1(y(ld-n))\beta_n + \alpha_2 g_2(y(ld-n))\beta_n$$
$$+\cdots+\alpha_m g_m(y(ld-n))\beta_n$$
$$= [g_1(y(ld-1)), g_1(y(ld-2)), \cdots,$$
$$g_1(y(ld-n))]\boldsymbol{\beta} + [g_2(y(ld-1))\beta_1,$$
$$g_2(y(ld-2))\beta_2, \cdots, g_2(y(ld-n))\beta_n]\alpha_2$$
$$+[g_3(y(ld-1))\beta_1, g_3(y(ld-2))\beta_2,$$
$$\cdots, g_3(y(ld-n))\beta_n]\alpha_3 + \cdots$$
$$+[g_m(y(ld-1))\beta_1, g_m(y(ld-2))\beta_2,$$
$$\cdots, g_m(y(ld-n))\beta_n]\alpha_m$$
$$= [\boldsymbol{g}_1(ld), \boldsymbol{g}_2(ld)\boldsymbol{\beta}, \boldsymbol{g}_3(ld)\boldsymbol{\beta}, \cdots,$$
$$\boldsymbol{g}_m(ld)\boldsymbol{\beta}] \begin{bmatrix} \boldsymbol{\beta} \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_m \end{bmatrix}$$
$$= \boldsymbol{\varphi}_2^{\mathrm{T}}(ld)\boldsymbol{\eta},$$

where

$$\boldsymbol{g}_i(ld) := [g_i(y(ld-1)), g_i(y(ld-2)), \cdots, g_i(y(1d-d)),$$
$$\cdots, g_i(y(ld-n))] \in \mathbb{R}^{1\times n},$$
$$i = 1, 2, \cdots, m,$$
$$\boldsymbol{\varphi}_2(ld) := [\boldsymbol{g}_1(ld), \boldsymbol{g}_2(ld)\boldsymbol{\beta}, \boldsymbol{g}_3(ld)\boldsymbol{\beta}, \cdots,$$
$$\boldsymbol{g}_m(ld)\boldsymbol{\beta}]^{\mathrm{T}} \in \mathbb{R}^{n+m-1},$$
$$\boldsymbol{\eta} := [\boldsymbol{\beta}, \alpha_2, \alpha_3, \cdots, \alpha_m]^{\mathrm{T}} \in \mathbb{R}^{n+m-1}.$$

Likewise, the dual-rate bilinear-in-parameter model in (3)–(5) becomes:

$$y(ld) = \boldsymbol{\varphi}_2^{\mathrm{T}}(ld)\boldsymbol{\eta} + \boldsymbol{\varphi}_u^{\mathrm{T}}(ld)\boldsymbol{\gamma} + v(ld)$$
$$= \boldsymbol{\psi}^{\mathrm{T}}(ld)\boldsymbol{\vartheta} + v(ld),$$

where

$$\boldsymbol{\psi}(ld) := \begin{bmatrix} \boldsymbol{\varphi}_2(ld) \\ \boldsymbol{\varphi}_u(ld) \end{bmatrix} \in \mathbb{R}^{n+m+n_\gamma-1},$$
$$\boldsymbol{\vartheta} := \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\gamma} \end{bmatrix} \in \mathbb{R}^{n+m+n_\gamma-1}.$$

Then employing the negative gradient search principle, the corresponding GI algorithm based on the parameter separation technique (the PS-GI algorithm for short) for the identification of system parameters can be written as

$$\hat{\boldsymbol{\vartheta}}_k = \hat{\boldsymbol{\vartheta}}_{k-1} + \mu_{2,k} \sum_{l=1}^{L} \hat{\boldsymbol{\psi}}_k(ld)[y(ld) - \hat{\boldsymbol{\psi}}_k^{\mathrm{T}}(ld)\hat{\boldsymbol{\vartheta}}_{k-1}], \tag{16}$$

$$\mu_{2,k} \le 2\lambda_{\max}^{-1} \left[\sum_{l=1}^{L} \hat{\boldsymbol{\psi}}_k(ld)\hat{\boldsymbol{\psi}}_k^{\mathrm{T}}(ld) \right], \tag{17}$$

$$\hat{\boldsymbol{\psi}}_k(ld) = [\hat{\boldsymbol{\varphi}}_{2,k}^{\mathrm{T}}(ld), \boldsymbol{\varphi}_u^{\mathrm{T}}(ld)]^{\mathrm{T}}, \tag{18}$$

$$\hat{\boldsymbol{\varphi}}_{2,k}(ld) = [\hat{\boldsymbol{g}}_{1,k}(ld), \hat{\boldsymbol{g}}_{2,k}(ld)\hat{\boldsymbol{\beta}}_{k-1}, \hat{\boldsymbol{g}}_{3,k}(ld)\hat{\boldsymbol{\beta}}_{k-1}, \cdots,$$
$$\hat{\boldsymbol{g}}_{m,k}(ld)\hat{\boldsymbol{\beta}}_{k-1}]^{\mathrm{T}}, \tag{19}$$

$$\hat{\boldsymbol{g}}_{i,k}(ld) = [g_i(\hat{y}_k(ld-1)), g_i(\hat{y}_k(ld-2)), \cdots,$$
$$g_i(y(ld-d)), \cdots, g_i(\hat{y}_k(ld-n))], \tag{20}$$

$$\boldsymbol{\varphi}_u(ld) = [u(ld-1), u(ld-2), \cdots, u(ld-n_\gamma)]^{\mathrm{T}}, \tag{21}$$

$$\hat{y}_k(ld-i) = \hat{\boldsymbol{\psi}}_k^{\mathrm{T}}(ld-i)\hat{\boldsymbol{\vartheta}}_k, \tag{22}$$

$$\hat{\boldsymbol{\vartheta}}_k = [\hat{\boldsymbol{\beta}}_k^{\mathrm{T}}, \hat{\alpha}_{2,k}, \hat{\alpha}_{3,k}, \cdots, \hat{\alpha}_{m,k}, \hat{\boldsymbol{\gamma}}_k^{\mathrm{T}}]^{\mathrm{T}}. \tag{23}$$

3 The particle filter-based GI algorithms

Data deficiency is commonly encountered in statistics, and it is important to address this in an appropriate way that does not interfere with the conclusions drawn from the given data. In this section, the objective is to derive the improved GI algorithms based on the particle filter which use the incomplete data to estimate concurrently the unknown system parameters and outputs. The particle filter is employed to calculate the unknown output $\hat{y}_k(ld-i)$, and the GI method is applied to estimate the value of the parameter vector $\boldsymbol{\vartheta}$.

To the best of our knowledge, one of attractive advantages of the particle filter lies in fact that it combines the Monte Carlo sampling methods with Bayesian inference to tackle the nonlinear dynamic estimation problem at a reasonable computational cost [65]. Hence, it is sometimes called the sequential Monte Carlo Bayesian estimator of the probability distribution of unknown variables. Within a Bayesian framework, it is assumed that the distribution of variables to be estimated is determined by all available variables in the system, which are generally considered to be random. In this regard, the first step is to represent the posterior probability density function $p(y(ld-i)|Y_o, U, \hat{\boldsymbol{\vartheta}}_k)$ based on the observed data $\{U, Y_o\}$ and the known parameter estimation vector $\hat{\boldsymbol{\vartheta}}_k$. Then, $\hat{y}_k(ld-i)$ is given by

$$\hat{y}_k(ld-i) = \int y(ld-i)f(y(ld-i))dy(ld-i),$$
$$f(y(ld-i)) := p(y(ld-i)|Y_o, U, \hat{\boldsymbol{\vartheta}}_k). \qquad (24)$$

Here, we denote $Y_o := \{y(d), y(2d), \cdots, y(ld)\}$ and $U := \{u(1), u(2), \cdots, u(ld)\}$, respectively. Note that typically this expression cannot be used directly to compute $\hat{y}_k(ld-i)$ analytically and, as a result, one has to resort to numerical techniques for its derivation. The particle filter achieves the task of representing the posterior probability density function by utilizing a bank of the independent random variables called particles. According to the Monte Carlo sampling principle, the right-hand side of (24) can be modified in the following form:

$$\int y(ld-i)f(y(ld-i))dy(ld-i) \approx \frac{1}{S}\sum_{s=1}^{S} y_k^{(s)}(ld-i),$$
$$= \frac{1}{S}\sum_{s=1}^{S}\int y(ld-i)\delta(y(ld-i)-\hat{y}_k^{(s)}(ld-i))dy(ld-i).$$

Therefore the posterior probability density function can be approximated as

$$f(y(ld-i)) \approx \frac{1}{S}\sum_{s=1}^{S}\delta(y(ld-i)-\hat{y}_k^{(s)}(ld-i)),$$

where S is the total number of samples, $\hat{y}_k^{(s)}(ld-i)$ is the s-th particle that approximates the distribution, and $\delta(\cdot)$ denotes the Dirac delta function. Then

$$\hat{y}_k(ld-i) \approx \frac{1}{S}\sum_{s=1}^{S}\hat{y}_k^{(s)}(ld-i).$$

However, it is not generally convenient to sample from the posterior probability density function directly; thus, we need to introduce the importance sampling technique. By defining $\hat{Y}_k(ld-i-1) := \{\hat{y}_k(ld-i-1), \cdots, y(ld-d), \cdots, \hat{y}_k(1)\}$, we choose the importance function $q(\cdot)$ as

$$q(y(ld-i)|Y(d:ld), U, \hat{\boldsymbol{\vartheta}}_k)$$
$$= p(y(ld-i)|\hat{Y}_k(ld-i-1), U, \hat{\boldsymbol{\vartheta}}_k),$$

which is in practice easier to sample than the target distribution.

In the importance sampling process, each sample is extracted from the importance distribution $q(\cdot)$ instead of the posterior $p(\cdot)$. This discrepancy is generally compensated by appropriately weighting each sample separately. Specifically, the associated weight $\hat{w}_k^{(s)}(ld)$ corresponding to $\hat{y}_k^{(s)}(ld)$ is determined by

$$\hat{w}_k^{(s)}(ld) \propto p(y(ld)|\hat{y}_k^{(s)}(ld))\hat{w}_k^{(s)}(ld-d). \qquad (25)$$

The posterior probability density function can then be represented by

$$f(y(ld-i)) \approx \sum_{s}^{S}\hat{W}_k^{(s)}(ld-i)\delta(y(ld-i)$$
$$-\hat{y}_k^{(s)}(ld-i)), \qquad (26)$$
$$\hat{W}_k^{(s)}(ld-i) = \frac{\hat{w}_k^{(s)}(ld-i)}{\sum_{s=1}^{S}\hat{w}_k^{(s)}(ld-i)}, \quad \sum_{s}^{S}\hat{W}_k^{(s)} = 1,$$

where $\hat{w}_k^{(s)}(ld-i)$ is the unnormalized importance weight and $\hat{W}_k^{(s)}(ld-i)$ is the normalized weight associated with the s-th particle at time $ld-i$.

Nevertheless, the probability density function $p(y(ld)|\hat{y}_k^{(s)}(ld))$ involved is still unknown. Assume that the noise is of Gaussian distribution. Then the weight can be represented by

$$\hat{w}_k^{(s)}(ld) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(y(ld)-\hat{y}_k^{(s)}(ld))^2}{2\sigma^2}\right]$$
$$\times \hat{w}_k^{(s)}(ld-d).$$

In what follows, we design a kernel density estimator to update $\hat{w}_k^{(s)}(ld)$ recursively, which offers a flexible way of modeling in the given datasets without the imposition of a parametric model. The kernel is a weighting function used in nonparametric identification methods. In order to construct the kernel function, we define

$$\hat{\xi}_k^{(s)}(ld) := |y(ld)-\hat{y}_k^{(s)}(ld)| > 0, \qquad (27)$$
$$\hat{\xi}_k(ld) := \max\{\hat{\xi}_k^{(s)}(ld), \quad s=1,2,\cdots,S\}+1. \qquad (28)$$

Clearly, we can obtain the following relationship:

$$\hat{\pi}_k^{(s)}(ld) := \frac{\hat{\xi}_k^{(s)}(ld)}{\hat{\xi}_k(ld)} < 1, \quad s=1,2,\cdots,S,$$

in which $\hat{\pi}_k^{(s)}(ld)$ can be viewed as the independent and identically distributed sample drawn from $p(y(ld)|\hat{y}_k^{(s)}(ld))$. Here, we choose the Epanechnikov function [66] to estimate $p(y(ld)|\hat{y}_k^{(s)}(ld))$. Then we obtain

$$p(y(ld)|\hat{y}_k^{(s)}(ld)) = \frac{3}{4}[1-(\hat{\pi}_k^{(s)}(ld))^2].$$

Upon normalizing the density function, one can get

$$p(y(ld)|\hat{y}_k^{(s)}(ld)) = \frac{\frac{3}{4}[1 - (\hat{\pi}_k^{(s)}(ld))^2]}{\sum_{s=1}^{S} \frac{3}{4}[1 - (\hat{\pi}_k^{(s)}(ld))^2]}$$

$$= \frac{1 - (\hat{\pi}_k^{(s)}(ld))^2}{S - \sum_{s=1}^{S} (\hat{\pi}_k^{(s)}(ld))^2}. \qquad (29)$$

Normalization guarantees that the result of the resulting estimation is a probability density function. Substituting (29) into (25) yields the weight update law:

$$\hat{w}_k^{(s)}(ld) = \frac{1 - (\hat{\pi}_k^{(s)}(ld))^2}{S - \sum_{s=1}^{S}(\hat{\pi}_k^{(s)}(ld))^2} \hat{w}_k^{(s)}(ld - d). \quad (30)$$

Remark 2 Noted that the above nonparametric kernel density estimation method is not flexible because of the definition of $\hat{\xi}_k(ld)$. In general, the smaller $\hat{\xi}_k^{(s)}(ld)$ is, the greater the weight of $\hat{y}_k^{(s)}(ld)$ is in estimation $y(ld)$, which may cause $\hat{\xi}_k(ld) \approx 1$. Therefore one new flexible definition for $\hat{\xi}_k(ld)$ is given by

$$\hat{\xi}_k(ld) := \max\{\hat{\xi}_k^{(s)}(ld), \quad s = 1, 2, \cdots, S\} + \chi^{\epsilon}, \tag{31}$$

where χ and ϵ are the adaptive factors. The value of $\hat{\xi}_k(ld)$ can be modified appropriately according to the adaptive factors. Therefore the new adaptive kernel estimator can be adjusted to improve the accuracy in fitting density function.

From the previous expression, it can easily be checked that the weights $\hat{w}_k^{(s)}(ld - i), i \neq 0$ cannot be determined based on (30). Thus, we keep the weights unchanged at the inter-sampled instant $ld - i$, i.e.,

$$\hat{w}_k^{(s)}(ld - i) = \hat{w}_k^{(s)}(ld), \quad i = 1, 2, \cdots, d - 1. \quad (32)$$

Accordingly, the missing outputs sought can now be calculated by a series of the weighted samples as follows:

$$\hat{y}_k(ld - i) = \sum_{s=1}^{S} \hat{W}_k^{(s)}(ld - i)\hat{y}_k^{(s)}(ld - i)$$

$$= \sum_{s=1}^{S} \hat{W}_k^{(s)}(ld)\hat{y}_k^{(s)}(ld - i).$$

In the particle filtering approach, resampling procedures are usually applied to reduce the adverse effects caused by the particle degeneracy problem. In this paper, we adopt the approach to filter the particles according to $\hat{W}_k^{(s)}(ld)$, i.e., discarding the samples with small weights and copying those with large weights,

before the weight of each sample is finally assigned as $\hat{W}_k^{(s)}(ld - i) = 1/S$. To be specific, the posterior probability density function and the missing output estimate are given by

$$f(y(ld - i)) \approx \sum_{r=1}^{S} \frac{1}{S}\delta(y(ld - i) - \hat{y}_k^{(r)}(ld - i)) \tag{33}$$

$$= \sum_{s=1}^{S} \frac{s_i}{S}\delta(y(ld - i) - \hat{y}_k^{(s)}(ld - i)),$$

$$\hat{y}_k(ld - i) \approx \sum_{r=1}^{S} \frac{1}{S}\hat{y}_k^{(r)}(ld - i) \tag{34}$$

$$= \sum_{s=1}^{S} \frac{s_i}{S}\hat{y}_k^{(s)}(ld - i),$$

$$\hat{y}_k^{(s)}(ld - i) \sim p(y(ld - i)|\hat{y}_k(ld - i - 1), \cdots,$$
$$\cdots \hat{y}_k(1), u(1), \cdots, u(Ld), \hat{\boldsymbol{\vartheta}}_k), \tag{35}$$

where $\hat{y}_k^{(r)}(ld - i)$ represents the r-th particle in a new set of particles resampled at time $ld - i$ and s_i refers to the number of times that $\hat{y}_k^{(s)}(ld - i)$ is copied in this new particle swarm. Equations (27)–(35) constitute the particle filtering algorithm. In the O-GI and PS-GI algorithms, we use (12) and (22) to calculate the output estimate $\hat{y}_k(ld - i)$, respectively. In the improved GI algorithms, we use the particle filter to estimate $y(ld - i)$. To be specific, replacing $\hat{y}_k(ld-i)$ in (12) and (22) with (34) yields a particle filter gradient-based iterative algorithm by using the over-parameterization technique (the O-PF-GI algorithm for short) and a particle filter gradient-based iterative algorithm by using the parameter separation technique (the PS-PF-GI algorithm for short), respectively. For simplicity, the pseudo-code description of the PS-PF-GI algorithm is provided in Algorithm 1.

Remark 3 The particle filter is an optimal estimation tool for nonlinear state-space models with unknown states, which represents the posterior distribution through a set of random particles and their associated weights. Inspired by the idea, a new particle filter-based algorithm for identifying a class of input-output systems within the GI framework is presented. The unknown outputs are regarded as the unknown states, while the known outputs are regarded as the measurable outputs. The kernel density estimation method is then developed

to update the weights of these particles at each iteration. Compared with the traditional GI method, the advantage of this method is that it can more accurately estimate the missing outputs. It is of the fact that the novel particle filter applies the measurable outputs to adjust the weight of each particle. In addition, the proposed algorithms in the paper can find many potential applications in the recovery and reconstruction of measurement signals in the noisy environment. The proposed particle filter-based algorithm of simultaneous output and parameter estimation for output nonlinear systems under low measurement rate constraints in this paper can be extended to multi-input and multi-output bilinear systems, and engineering technology literature [67–70] can be applied to other control and schedule areas such as the information processing and transportation communication systems [71–77] and so on.

4 Example

In this section, we provide a detailed numerical experiment to evaluate the ability for parameter estimation and numerical performance in terms of accuracy and efficiency in the applications of the four proposed methods, namely the PS-PF-GI, the O-PF-GI, the PS-GI and the O-GI algorithms. This comparison will hence allow us to see how the proposed the particle filter-based approaches perform compared to the traditional solutions. The system under consideration is given by

$$y(l) = [1, 0.28] \begin{bmatrix} 0.30y(l-1) & 0.30y(l-2) & 0.30y(l-3) \\ 0.50y^2(l-1) & 0.50y^2(l-2) & 0.50y^2(l-3) \end{bmatrix}$$

$$\begin{bmatrix} 0.45 \\ 0.30 \\ 0.46 \end{bmatrix} + 0.10u(l-1) + 0.90u(l-2) + v(l),$$

$$\boldsymbol{\alpha} = [1, 0.28]^{\mathrm{T}}, \quad \boldsymbol{\beta} = [0.45, 0.30, 0.46]^{\mathrm{T}},$$
$$\boldsymbol{\gamma} = [0.10, 0.90]^{\mathrm{T}},$$
$$\boldsymbol{\vartheta} = [0.28, 0.45, 0.30, 0.46, 0.10, 0.90]^{\mathrm{T}}.$$

In simulation, the input $u(l)$ is taken as the persistent excitation signal with zero mean and unit variance and $v(l)$ as a white noise with zero mean. The number of the particles in the particle filter is set to be $S = 50$, and the measurement data length is taken as $L = 600$. The first 400 data samples ($l = 1, 2, 3, \cdots, 400$) are used in the first stage to estimate $\boldsymbol{\vartheta}$, $y(l-i)$, and the remaining data points (samples 400-600) are then used to test the estimated models. When taking $d = 2$, the inputs $\{u(1), u(2), u(3), \cdots, u(400)\}$

Algorithm 1 The PS-PF-GI algorithm

Require: $u(1), u(2), \cdots, u(Ld), y(d), y(2d), \cdots, y(Ld)$
    ($n_0 := m + n + n_\gamma - 1, L \gg n_0$)
Ensure: Set $\hat{\boldsymbol{\vartheta}}_0 = \mathbf{1}_{n_0}/p0$, $\hat{y}_1(ld - i) = 1/p0$, $p0 = 10^6$.
        Give the nonlinear function $g_i(\cdot)$ and a small positive
    number κ.
        Draw the initial S particles $\{\hat{y}_1^{(s)}(d - i)\}_{s=1}^S$ and let
  $\{\hat{W}_1^{(s)}(0)\}_{s=1}^S = \frac{1}{S}$.
1: **Estimate** $\hat{\boldsymbol{\vartheta}}_k$
2: **for** $k = 1, 2, \cdots$ **do**
3:   Form $\hat{\boldsymbol{\phi}}_{2,k}(ld)$ using (19) for $l = 1, 2, \cdots, L$.
4:   Form $\hat{\boldsymbol{\psi}}_k(ld)$ using (18) for $l = 1, 2, \cdots, L$.
5:   Compute $\mu_{2,k}(l)$ using (17).
6:   Update $\hat{\boldsymbol{\vartheta}}_k$ using (16).
7:   **Update** $\hat{y}_k(ld - i)$, $i = 1, 2, \cdots, d - 1$, $l = 1, 2, \cdots, L$.
8:   **for** $l = 1 : L$ **do**
9:     **while** $i = 0$ **do**
10:       Sample $\{\hat{y}_k^{(s)}(ld)\}_{s=1}^S$ from $p(y(ld)|\hat{Y}_k(ld-1), U, \hat{\boldsymbol{\vartheta}}_k)$.
11:       Calculate $\hat{\xi}_k^{(s)}(ld)$, for $s = 1, 2, \cdots, S$ and obtain $\hat{\xi}_k(ld)$.
12:       Calculate $\hat{\pi}_k^{(s)}(ld) = \frac{\hat{\xi}_k^{(s)}(ld)}{\hat{\xi}_k(ld)}$.
13:       Calculate $\hat{w}_k^{(s)}(ld) = \frac{1-(\hat{\pi}_k^{(s)}(ld))^2}{S-\sum_{s=1}^S (\hat{\pi}_k^{(s)}(ld))^2} \hat{w}_k^{(s)}(ld-d)$.
14:       Normalize the weight $\{\hat{W}_k^{(s)}(ld)\}_s^S = \frac{\hat{w}_k^{(s)}(ld)}{\sum_{s=1}^S \hat{w}_k^{(s)}(ld)}$.
15:     **end while**
16:     **for** $i = d - 1 : 1$ **do**
17:       Sample $\{\hat{y}_k^{(s)}(ld - i)\}_{s=1}^S$ from $p(y(ld - i)|\hat{Y}_k(ld - i - 1), U, \hat{\boldsymbol{\vartheta}}_k)$.
18:       Resample the new particle set $\{\hat{y}_k^{(r)}(ld - i)\}_r^S$:
19:         select $\{\hat{y}_k^{(r)}(ld - i)\}_{r=1}^S$ from $\{\hat{y}_k^{(s)}(ld - i)\}_{s=1}^S$ according to $\hat{W}_k^{(s)}(ld)$.
20:       Calculate $\hat{y}_k(ld - i) = \frac{1}{S} \sum_{r=1}^S \hat{y}_k^{(r)}(ld - i)\}_r^S$.
21:     **end for**
22:   **end for**
23:   **if** $\|\hat{\boldsymbol{\vartheta}}_k - \hat{\boldsymbol{\vartheta}}_{k-1}\| > \kappa$ **then**
24:     $k = k + 1$
25:   **else**
26:   Obtain the iteration k and the parameter estimation vector $\hat{\boldsymbol{\vartheta}}_k$, break;
27:   **end if**
28: **end for**

---

and the outputs $\{y(2), y(4), \cdots, y(400)\}$ are available, while $\{y(1), y(3), \cdots, y(399)\}$ are unavailable.

The simulation consists of three parts: (i) the unknown output estimation and (ii) the system parameter estimation and (iii) the output prediction. The results are given in Tables 1, 2 and Figs. 2, 3, 4, 5, and 6, respectively. Table 1 shows the parameter estimates and the corresponding parameter estimation errors $\delta := \|\hat{\boldsymbol{\vartheta}}_k(l) - \boldsymbol{\vartheta}\|/\|\boldsymbol{\vartheta}\|$ obtained by the four

Table 1 The parameter estimates and their errors under different algorithms

| Algorithms	Variance σ^2	α_2	β_1	β_2	β_3	γ_1	γ_2	$\delta(\%)$
PS-PF-GI	0.10^2	0.28302	0.44848	0.29918	0.44996	0.09503	0.90193	1.00429
	0.30^2	0.28612	0.43999	0.30177	0.43238	0.08293	0.90834	3.00284
	0.50^2	0.28482	0.42704	0.31578	0.42032	0.06899	0.91730	5.09108
	0.70^2	0.27395	0.41711	0.34647	0.42137	0.05339	0.92967	7.46843
	0.90^2	0.23858	0.42080	0.40059	0.44461	0.03890	0.94673	11.59058
PS-GI	0.10^2	0.28375	0.44388	0.31406	0.43739	0.09271	0.90880	2.52062
	0.30^2	0.28418	0.44147	0.31341	0.42043	0.08009	0.91599	4.21179
	0.50^2	0.27988	0.43389	0.32432	0.40906	0.06553	0.92567	6.14361
	0.70^2	0.26551	0.42817	0.35269	0.41024	0.04961	0.93903	8.44468
	0.90^2	0.23416	0.43509	0.40635	0.43472	0.03367	0.95795	12.53349
O-PF-GI	0.10^2	0.29261	0.44462	0.28841	0.46067	0.09684	0.90325	1.56486
	0.30^2	0.29320	0.42160	0.29636	0.45296	0.08575	0.91145	3.13413
	0.50^2	0.28221	0.38341	0.34449	0.43488	0.07169	0.92199	7.70822
	0.70^2	0.28460	0.33412	0.41945	0.42573	0.05687	0.93585	15.11413
	0.90^2	0.29301	0.28156	0.51598	0.44645	0.04277	0.95390	24.11411
O-GI	0.10^2	0.29755	0.43119	0.31334	0.43288	0.09574	0.90846	3.44552
	0.30^2	0.29900	0.41361	0.31932	0.41588	0.08413	0.91603	5.67330
	0.50^2	0.29387	0.38083	0.36355	0.38969	0.06976	0.92584	10.52666
	0.70^2	0.29951	0.33637	0.43389	0.37568	0.05477	0.93911	17.28523
	0.90^2	0.30439	0.28725	0.52671	0.39673	0.04054	0.95714	25.23102
True values		0.28000	0.45000	0.30000	0.46000	0.10000	0.90000	

Table 2 The means and the standard deviations of the output estimation error under different algorithms

Algorithms	Variance σ^2	Means	Stds	Variance σ^2	Means	Stds
PS-PF-GI	0.10^2	-0.00109	0.10158	0.30^2	0.00457	0.28753
	0.50^2	0.01622	0.47568	0.70^2	0.03319	0.66526
	0.90^2	0.06517	0.85807	1.00^2	0.13071	0.95817
PS-GI	0.10^2	0.01328	0.15557	0.30^2	0.01975	0.31034
	0.50^2	0.03299	0.49045	0.70^2	0.05346	0.67685
	0.90^2	0.08131	0.86188	1.00^2	0.09627	0.96635
O-PF-GI	0.10^2	-0.00160	0.09746	0.30^2	0.00743	0.28578
	0.50^2	0.02395	0.47695	0.70^2	0.04273	0.66892
	0.90^2	0.06331	0.86189	1.00^2	0.07293	0.95927
O-GI	0.10^2	0.00858	0.13787	0.30^2	0.01735	0.30325
	0.50^2	0.03289	0.48799	0.70^2	0.05074	0.67653
	0.90^2	0.07066	0.86743	1.00^2	0.08585	0.96431

Fig. 2 The PS-PF-GI estimation errors δ versus k
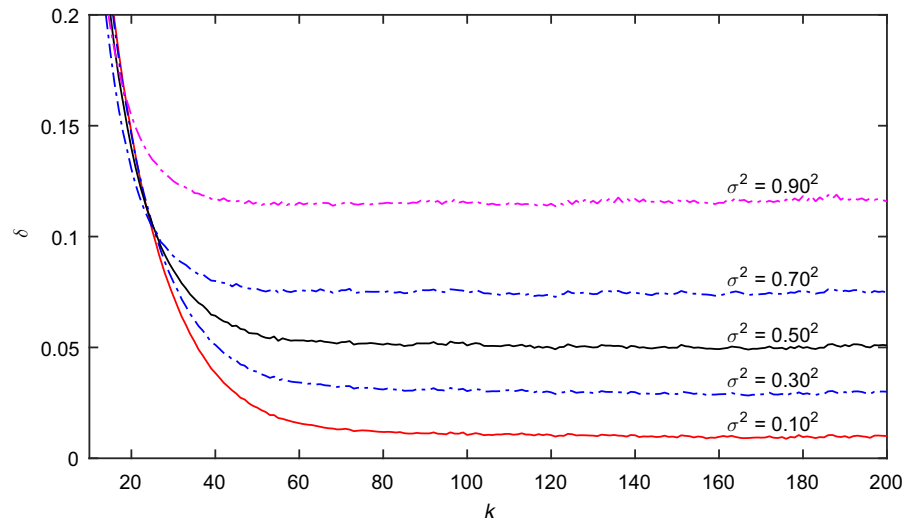
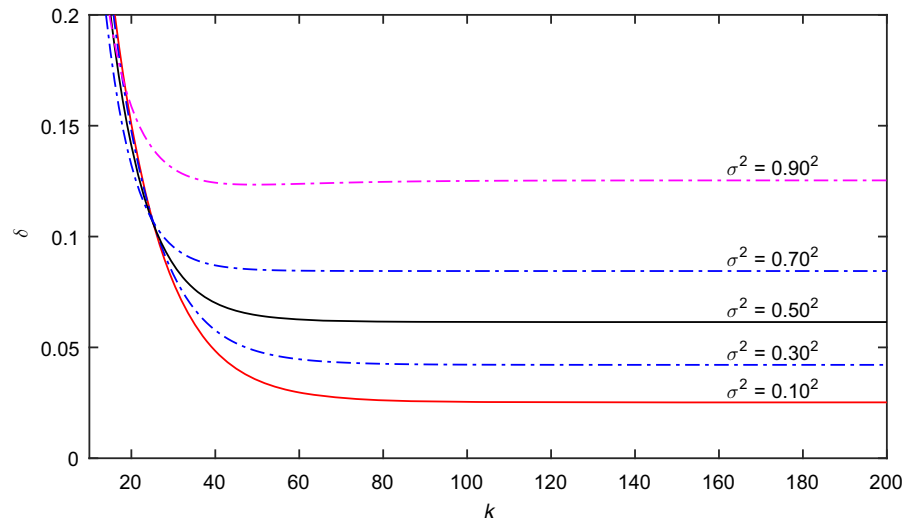**Fig. 3** The O-PF-GI estimation errors δ versus k
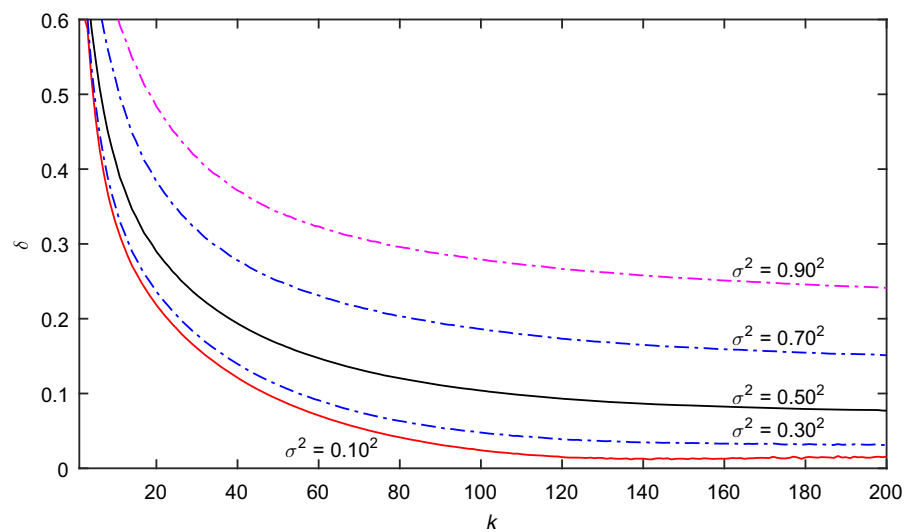
algorithms under the different noise variances when the number of iterations $k = 200$. Then, the relationships between δ and k are plotted in Figs. 2, 3, 4 and 5. The means and standard deviations (Stds) of the errors $\epsilon := y(l) - \hat{y}_k(l)$ between the true missing outputs and the estimated outputs under the same conditions are shown in Table 2. To test the quality of the estimated models obtained from the different algorithms, we compare the errors between the predicted outputs and the true outputs by using $l_e = 200$ samples from $l = 401$ to $l = 600$. The detailed comparison results are shown in Fig. 6, where the iteration $k = 200$ and the noise variance $\sigma^2 = 0.10^2$. Using the root-mean-square error to denote the errors, we have

$$Error_1 = \left[\frac{1}{l_e} \sum_{l=401}^{600} [\hat{y}_{1,k}(l) - y(l)]^2 \right]^{1/2} = 0.10288,$$

$$Error_2 = \left[\frac{1}{l_e} \sum_{l=401}^{600} [\hat{y}_{2,k}(l) - y(l)]^2 \right]^{1/2} = 0.10308,$$

$$Error_3 = \left[\frac{1}{l_e} \sum_{l=401}^{600} [\hat{y}_{3,k}(l) - y(l)]^2 \right]^{1/2} = 0.10446,$$

$$Error_4 = \left[\frac{1}{l_e} \sum_{l=401}^{600} [\hat{y}_{4,k}(l) - y(l)]^2 \right]^{1/2} = 0.10467,$$

where $\hat{y}_{1,k}(l)$ denotes the predicted output of the PS-PF-GI algorithm, $\hat{y}_{2,k}(l)$ the predicted output of the

Fig. 4 The PS-GI estimation errors δ versus k

**Fig. 5** The O-GI estimation errors δ versus k
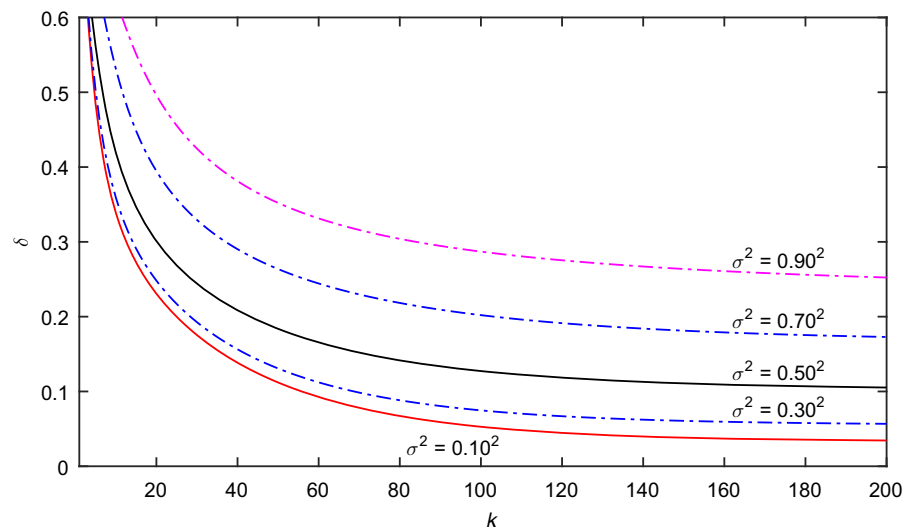
O-PF-GI algorithm, $\hat{y}_{3,k}(l)$ the predicted output of the PS-GI algorithm and $\hat{y}_{4,k}(l)$ the predicted output of the O-GI algorithm.

From the result of this experiment, we can draw the following features.

– It can be observed from the results in Table 1 that the particle filter-based algorithms generate lower estimation errors than the errors caused by the traditional methods under the different noise levels—see the parameter estimation errors in the last column in Table 1.

– According to Figs. 2, 3, 4 and 5, we can see clearly that the parameter estimation accuracy given by the PS-PF-GI algorithm, the PS-O-GI algorithm, the PS-GI algorithm and the O-GI algorithm becomes high with the iterative step k increasing. Under the same conditions, an appropriate noise level leads to lower parameter estimation accuracy.

– The estimated outputs of the PS-PF-GI algorithm can track the missing outputs very well, as shown in Table 2. It can be observed from Fig. 6 that the predicted outputs from the PS-PF-GI estimated model fluctuate in a small range around their true values.
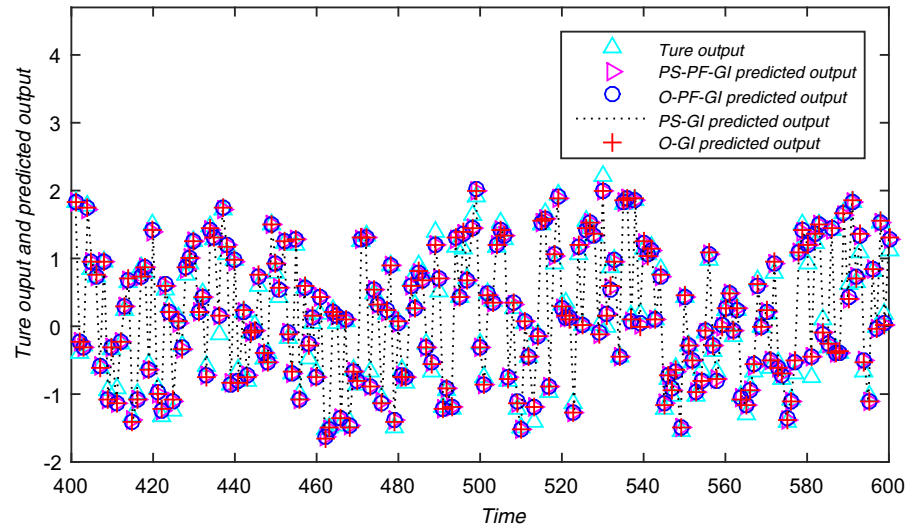
However, compared with the PS-GI and O-GI algorithms, the PS-PF-GI and O-PF-GI algorithms have relatively larger computational burden.

5 Conclusions

This paper focuses on a class of dual-rate nonlinear systems which have been widely used in nonlinear system modeling and control. A key issue is to effectively identify the systems with good generalization performance. In this regard, a combined particle-based filter and the gradient-based iterative method is presented to identify a dual-rate bilinear-in-parameter system. It involves two steps, where one estimates the joint probability density of the missing observations based on the initial guesses of the parameters in the first step, and the second step minimizes the cost function to update a new estimate of the parameter vector. These two steps are repeated until the change in the parameter estimates after each iteration is within a specified tolerance level. The numerical results of the simulation example indicate that the particle filter-based methods can give more reliable output and parameter estimation than the traditional gradient-based iterative algorithms. The methods proposed in this paper can combine some mathematical methods [78–85] and other identification approaches [86–92] to study the parameter estimation problems of other linear systems and nonlinear systems [93–99] and can be applied to other engineering systems.

Declarations

References

1. Ding, F., Liu, G., Liu, X.P.: Parameter estimation with scarce measurements. Automatica **47**(8), 1646–1655 (2011)
2. Ding, F., Liu, Y.J., Bao, B.: Gradient based and least squares based iterative estimation algorithms for multi-input multi-output systems. Proc. Inst. Mech. Eng. Part I: J. Syst. Control Eng. **226**(1), 43–55 (2012)
3. Zhang, K.W., Shi, Y.: Adaptive model predictive control for a class of constrained linear systems with parametric uncertainties. Automatica **117**, 108974 (2020)
4. Ding, F.: System Identification - Performances Analysis for Identification Methods. Science Press, Beijing (2014)
5. Stojanovic, V., Prsic, D.: Robust identification for fault detection in the presence of non-Gaussian noises: application to hydraulic servo drives. Nonlinear Dyn. **100**(3), 2299–2313 (2020)
6. Lin, R.M., Ng, T.Y.: A new method for the accurate measurement of higher-order frequency response functions of nonlinear structural systems. ISA Trans. **81**, 270–285 (2018)
7. Xu, L., Song, G.L.: A recursive parameter estimation algorithm for modeling signals with multi-frequencies. Circuits Syst. Signal Process. **39**(8), 4198–4224 (2020)
8. Gan, M., Chen, G., Chen, L., Chen, C.L.P.: Term selection for a class of separable nonlinear models. IEEE Trans. Neural Netw. Learn. Syst. **31**(2), 445–451 (2020)

9. Lin, R.M., Ng, T.Y.: Higher-order FRFs and their applications to the identifications of continuous structural systems with discrete localized nonlinearities. Mech. Syst. Signal Proc. **108**, 326–346 (2018)

10. Zhu, Q.M., Zhang, W.C., Zhang, J.H., Sun, B.: U-neural network-enhanced control of nonlinear dynamic systems. Neurocomputing **352**, 12–21 (2019)

11. Wang, L.J., Ji, Y., Wan, L.J., Bu, N.: Hierarchical recursive generalized extended least squares estimation algorithms for a class of nonlinear stochastic systems with colored noise. J. Frankl. Inst. **356**(16), 10102–10122 (2019)

12. Ren, H.R., Karimi, H.R., Lu, R.Q., Wu, Y.Q.: Synchronization of network systems via aperiodic sampled-data control with constant delay and application to unmanned ground vehicles. IEEE Trans. Ind. Electron. **67**(6), 4980–4990 (2020)

13. Allen, J., Chen, S.R., El-Farra, N.H.: Model-based strategies for sensor fault accommodation in uncertain dynamic processes with multi-rate sampled measurements. Chem. Eng. Res. Des. **142**, 204–213 (2019)

14. Sanjurjo, E., Naya, M.A., Blanco-Claraco, J.L., Torres-Moreno, J.L., Gimenez-Fernandez, A.: Accuracy and efficiency comparison of various nonlinear Kalman filters applied to multibody models. Nonlinear Dyn. **88**(3), 1935–1951 (2017)

15. Cuenca, A., Zheng, M.H., Tomizuka, M., Sanchez, S.: Non-uniform multi-rate estimator based periodic event-triggered control for resource saving. Inf. Sci. **459**, 86–102 (2019)

16. Pasand, M.M.S., Montazeri, M.: Controllability and stabilizability of multi-rate sampled data systems. Syst. Control Lett. **113**, 27–30 (2018)

17. Li, J.H., Zhang, J.L.: Maximum likelihood identification of dual-rate Hammerstein output-error moving average system. IET Control Theory Appl. **14**(8), 1078–1090 (2020)

18. Yan, W.L., Du, C.L., Pang, C.K.: A general multirate approach for direct closed-loop identification to the Nyquist frequency and beyond. Automatica **53**, 164–170 (2015)

19. Chen, W.H., He, H.H., Lu, X.M.: Multi-rate sampled data composite control of linear singularly perturbed systems. J. Frankl. Inst. **357**(4), 2028–2048 (2020)

20. Yu, C.P., Chen, J., Verhaegen, M.: Subspace identification of individual systems in a large-scale heterogeneous network. Automatica **109**, 108517 (2019)

21. Guo, C.Q., Wang, L.J., Deng, F.: The auxiliary model based hierarchical estimation algorithms for bilinear stochastic systems with colored noises. Int. J. Control Autom. Syst. **18**(3), 650–660 (2020)

22. Ji, Y., Jiang, X.K., Wan, L.J.: Hierarchical least squares parameter estimation algorithm for two-input Hammerstein finite impulse response systems. J. Frankl. Inst. **357**(8), 5019–5032 (2020)

23. Li, M.H., Liu, X.M.: Maximum likelihood least squares based iterative estimation for a class of bilinear systems using the data filtering technique. Int. J. Control Autom. Syst. **18**(6), 1581–1592 (2020)

24. Li, M.H., Liu, X.M.: The filtering-based maximum likelihood iterative estimation algorithms for a special class of nonlinear systems with autoregressive moving average noise using the hierarchical identification principle. Int. J. Adapt. Control Signal Process. **33**(7), 1189–1211 (2019)

25. Li, M.H., Liu, X.M.: Maximum likelihood hierarchical least squares-based iterative identification for dual-rate stochastic systems. Int. J. Adapt. Control Signal Process. **35**(2), 240–261 (2021)

26. Li, M.H., Liu, X.M.: The least squares based iterative algorithms for parameter estimation of a bilinear system with autoregressive noise using the data filtering technique. Signal Process. **147**, 23–34 (2018)

27. Li, M.H., Liu, X.M.: Iterative parameter estimation methods for dual-rate sampled-data bilinear systems by means of the data filtering technique. IET Control Theory Appl. **15**(9), 1230–1245 (2021)

28. Fan, Y.M., Liu, X.M.: Two-stage auxiliary model gradient-based iterative algorithm for the input nonlinear controlled autoregressive system with variable-gain nonlinearity. Int. J. Robust Nonlinear Control **30**(14), 5492–5509 (2020)

29. Liu, X.M., Fan, Y.M.: Maximum likelihood extended gradient-based estimation algorithms for the input nonlinear controlled autoregressive moving average system with variable-gain nonlinearity. Int. J. Robust Nonlinear Control **31**(9), 4017–4036 (2021)

30. Ding, F.: System Identification—Iterative Search Principle and Identification Methods. Science Press, Beijing (2018)

31. Ji, Y., Zhang, C., Kang, Z., Yu, T.: Parameter estimation for block-oriented nonlinear systems using the key term separation. Int. J. Robust Nonlinear Control **30**(9), 3727–3752 (2020)

32. Zhao, Z.G., Tulsyan, A., Huang, B., Liu, F.: Estimation and identification in batch processes with particle filters. J. Process Control **81**, 1–14 (2019)

33. Lim, J.: Particle filtering for nonlinear dynamic state systems with unknown noise statistics. Nonlinear Dyn. **78**(2), 1369–1388 (2014)

34. Liu, X., Yang, X.Q., Zhu, P.B., Xiong, W.L.: Robust identification of nonlinear time-delay system in state-space form. J. Frankl. Inst. **356**(16), 9953–9971 (2019)

35. Risuleo, R.S., Lindsten, F., Hjalmarsson, H.: Bayesian nonparametric identification of Wiener systems. Automatica **108**, 108480 (2019)

36. Beeson, R., Namachchivaya, N.S.: Particle filtering for chaotic dynamical systems using future right-singular vectors. Nonlinear Dyn. **102**(2), 679–696 (2020)

37. Chen, J., Li, J., Liu, Y.J.: Gradient iterative algorithm for dual-rate nonlinear systems based on novel particle filter. J. Frankl. Inst. **354**(11), 4425–4437 (2017)

38. Ding, J., Chen, J.Z., Lin, J.X., Wan, L.J.: Particle filtering based parameter estimation for systems with output-error type model structures. J. Frankl. Inst. **356**(10), 5521–5540 (2019)

39. Chen, M.T.: Gradient-based iterative parameter estimation for bilinear-in-parameter systems using the model decomposition technique. IET Control Theory Appl. **12**(17), 2380–2389 (2018)

40. Chen, M.T.: Iterative identification algorithms for bilinear-in-parameter systems with autoregressive moving average noise. J. Frankl. Inst. **354**(17), 7885–7898 (2017)

41. Abrahamsson, R., Kay, S.M., Stoica, P.: Estimation of the parameters of a bilinear model with applications to submarine detection and system identification. Digit. Signal Process. **17**(4), 756–773 (2007)

42. Wang, L.J., Ji, Y., Yang, H.L., Xu, L.: Decomposition-based multiinnovation gradient identification algorithms for a special bilinear system based on its input-output representation. Int. J. Robust Nonlinear Control **30**(9), 3607–3623 (2020)

43. Paleologu, C., Benesty, J., Ciochină, S.: Adaptive filtering for the identification of bilinear forms. Digit. Signal Process. **75**, 153–167 (2018)

44. Gu, Y., Zhu, Q., Nouri, H.: Bias compensation-based parameter and state estimation for a class of time-delay nonlinear state-space models. IET Control Theory Appl. **14**(15), 2176–2185 (2020)

45. Gu, Y., Zhu, Q.M., Li, C.J., Zhun P,Y, Nouri, H.: State filtering and parameter estimation for two-input two-output systems with time delay. IET Control Theory Appl. **15**(16), 2053–2066, (2021).

46. Gu, Y., Chou, J., Liu, J., Ji, Y.: Moving horizon estimation for multirate systems with time-varying time-delays. J. Frankl. Inst **356**(4), 2325–2345 (2019)

47. Xu, L.: Application of the Newton iteration algorithm to the parameter estimation for dynamical systems. J. Comput. Appl. Math. **288**, 33–43 (2015)

48. Xu, L., Chen, L., Xiong, W.L.: Parameter estimation and controller design for dynamic systems from the step responses based on the Newton iteration. Nonlinear Dyn. **79**(3), 2155–2163 (2015)

49. Ji, Y., Kang, Z.: Three-stage forgetting factor stochastic gradient parameter estimation methods for a class of nonlinear systems. Int. J. Robust Nonlinear Control **31**(3), 871–987 (2021)

50. Xu, L.: The damping iterative parameter identification method for dynamical systems based on the sine signal measurement. Signal Process. **120**, 660–667 (2016)

51. Xu, L., Xiong, W.L., Alsaedi, A., Hayat, T.: Hierarchical parameter estimation for the frequency response based on the dynamical window data. Int. J. Control Autom. Syst. **16**(4), 1756–1764 (2018)

52. Ding, J., Cao, Z.X., Chen, J.Z., Jiang, G.P.: Weighted parameter estimation for Hammerstein nonlinear ARX systems. Circuits Syst. Signal Process. **39**(4), 2178–2192 (2020)

53. Zhang, X.: Highly computationally efficient state filter based on the delta operator. Int. J. Adapt. Control Signal Process. **33**(6), 875–889 (2019)

54. Zhang, X.: State estimation for bilinear systems through minimizing the covariance matrix of the state estimation errors. Int. J. Adapt. Control Signal Process. **33**(7), 1157–1173 (2019)

55. Wang, X.G., Wan, Z.W., Tang, L., et al.: Electromagnetic performance analysis of an axial flux hybrid excitation motor for HEV drives. IEEE Trans. Appl. Superconduct. **31**(8), 5205605 (2021)

56. Zhao, Z.Y., Wang, X.Y., Yao, P., Bai, Y.T.: A health performance evaluation method of multirotors under wind turbulence. Nonlinear Dyn. **102**(2), 1701–1715 (2020)

57. Ding, F., Wang, F.F., Xu, L., et al.: Parameter estimation for pseudo-linear systems using the auxiliary model and the decomposition technique. IET Control Theory Appl. **11**(3), 390–400 (2017)

58. Wang, X.G., Zhao, M., Zhou, Y., et al.: Design and analysis for multi-disc coreless axial-flux permanent-magnet synchronous machine. IEEE Trans. Appl. Superconductivity **31**(8), 5203804 (2021)

59. Wu, M.H., Yue, H.H., Wang, J., et al.: Object detection based on RGC mask R-CNN. IET Image Process. **14**(8), 1502–1508 (2020)

60. Kong, J.L., Wang, H.X., Wang, X.Y., et al.: Multi-stream hybrid architecture based on cross-level fusion strategy for fine-grained crop species recognition in precision agriculture. Comput. Electron. Agr. **185**, 106134 (2021)

61. Ding, F., Chen, H.B., et al.: A hierarchical least squares identification algorithm for Hammerstein nonlinear systems using the key term separation. J. Frankl. Inst. **355**(8), 3737–3752 (2018)

62. Ding, F., Xu, L., Alsaadi, F.E., Hayat, T.: Iterative parameter identification for pseudo-linear systems with ARMA noise using the filtering technique. IET Control Theory Appl. **12**(7), 892–899 (2018)

63. Zheng, Y.Y., Kong, J.L., Jin, X.B., et al. CropDeep: The crop vision dataset for deep-learning-based classification and detection in precision agriculture. Sensors **19**(5), 1058 (2019)

64. Wang, L.J., Guo, J., Xu, C., et al.: Hybrid model predictive control strategy of supercapacitor energy storage system based on double active bridge. Energies **12**(11), 2134 (2019)

65. Moghaddasi, S.S., Faraji, N.: A hybrid algorithm based on particle filter and genetic algorithm for target tracking. Expert Syst. Appl. **147**, 113188 (2020)

66. Wand, M.P., Jones, M.C.: Kernel Smoothing. CRC Press, Boca Raton, FL (1994)

67. Ding, F., Xu, L., Zhu, Q.: Performance analysis of the generalised projection identification for time-varying systems. IET Control Theory Appl. **10**(18), 2506–2514 (2016)

68. Xu, L.: The parameter estimation algorithms based on the dynamical response measurement data. Adv. Mech. Eng. **9**(11), Article Number: 1687814017730003 (2017)

69. Zhao, G., Cao, T.H., Yang, Y.D., et al.: Optimal sizing of isolated microgrid containing photovoltaic/photothermal/wind/diesel/battery. Int. J. Photoenergy **2021**, 5566597 (2021)

70. Zhang, Y., Yan, Z., Zhou, C.C., Wu, T.Z.: Capacity allocation of HESS in micro-grid based on ABC algorithm. Int. J. Low-Carbon Technol. **15**(4), 496–505 (2020)

71. Cao, Y., Wen, J.K., Ma, L.C.: Tracking and collision avoidance of virtual coupling train control system. Alex. Eng. J. **60**(2), 2115–2125 (2021)

72. Cao, Y., Wen, J.K., Ma, L.C.: Tracking and collision avoidance of virtual coupling train control system. Fut. Gen. Comput. Syst. **120**, 76–90 (2021)

73. Su, S., Wang, X.K., Cao, Y., Yin, J.T.: An energy-efficient train operation approach by integrating the metro timetabling and eco-driving. IEEE Trans. Intell. Transp. Syst. **21**(10), 4252–4268 (2020)

74. Su, S., Tang, T., Xu, J., Cao, F., Wang, Y., H.: Design of running grades for energy-efficient train regulation: A case study for beijing yizhuang line. IEEE Intell. Transp. Syst. Mag **13**(2), 189–200 (2021)

75. Cao, Y., Wang, Z., Liu, F., Li, P., Xie, G.: Bio-inspired speed curve optimization and sliding mode tracking control for subway trains. IEEE Trans. Veh. Technol. **68**(7), 6331–6342 (2019)

76. Cao, Y., Sun, Y.K., Xie, G., Wen, T.: Fault diagnosis of train plug door based on a hybrid criterion for IMFs selection

and fractional wavelet package energy entropy. IEEE Trans. Veh. Technol. **68**(8), 7544–7551 (2019)

77. Zhang, X.M., Zhao, Z.Y., Wang, Z.Y., Wang, X.Y.: Fault detection and identification method for quadcopter based on airframe vibration signals. Sensors **21**(2), Article Number: 581 (2021)

78. Dong, H., Yin, C.C., Dai, H.S.: Spectrally negative Levy risk model under Erlangized barrier strategy. J. Comput. Appl. Math. **351**, 101–116 (2019)

79. Sha, X.Y., Xu, Z.S., Yin, C.C.: Elliptical distribution-based weight-determining method for ordered weighted averaging operators. Int. J. Intell. Syst. **34**(5), 858–877 (2019)

80. Yin, C.C., Wen, Y.Z.: An extension of Paulsen-Gjessing's risk model with stochastic return on investments. Insurance Math. Econ. **52**(3), 469–476 (2013)

81. Zhao, Y.X., Chen, P., Yang, H.L.: Optimal periodic dividend and capital injection problem for spectrally positive Levy processes. Insurance Math. Econ. **74**, 135–146 (2017)

82. Zhao, X.H., Dong, H., Dai, H.S.: On spectrally positive Levy risk processes with Parisian implementation delays in dividend payments. Stat. Probab. Lett. **140**, 176–184 (2018)

83. Zhao, Y.X., Yin, C.C.: The expected discounted penalty function under a renewal risk model with stochastic income. Appl. Math. Comput. **218**(10), 6144–6154 (2012)

84. Hou, J., Chen, F.W., Li, P.H., Zhu, Z.Q.: Gray-box parsimonious subspace identification of Hammerstein-type systems. IEEE Trans. Ind. Electron. **68**(10), 9941–9951 (2021)

85. Li, X.Y., Wang, H.L., Wu, B.Y.: A stable and efficient technique for linear boundary value problems by applying kernel functions. Appl. Numer. Math. **172**, 206–214 (2022)

86. Xu, L.: Hierarchical Newton and least squares iterative estimation algorithm for dynamic systems by transfer functions based on the impulse responses. Int. J. Syst. Sci. **50**(1), 141–151 (2019)

87. Xu, L.: Separable multi-innovation stochastic gradient estimation algorithm for the nonlinear dynamic responses of systems. Int. J. Adapt. Control Signal Process. **34**(7), 937–954 (2020)

88. Pan, J., Ma, H., Zhang, X., et al.: Recursive coupled projection algorithms for multivariable output-error-like systems with coloured noises. IET Signal Process. **14**(7), 455–466 (2020)

89. Ma, H., Pan, J., et al.: Partially-coupled least squares based iterative parameter estimation for multi-variable output-error-like autoregressive moving average systems. IET Control Theory Appl. **13**(18), 3040–3051 (2019)

90. Ma, H., Zhang, X., Liu, Q.Y., et al.: Partiallly-coupled gradient-based iterative algorithms for multivariable output-error-like systems with autoregressive moving average noises. IET Control Theory Appl. **14**(17), 2613–2627 (2020)

91. Chen, G.Y., Gan, M., Chen, C.L.P., Li, H.X.: Basis function matrix-based flexible coefficient autoregressive models: A framework for time series and nonlinear system modeling. IEEE Trans. Cybern. **51**(2), 614–623 (2021)

92. Gan, M., Chen, X.X., Chen, C.L.P.: Adaptive RBF-AR models based on multi-innovation least squares method. IEEE Signal Process. Lett. **26**(8), 1182–1186 (2019)

93. Chen, G.Y., Gan, M., Chen, C.L.P., Li, H.X.: A regularized variable projection algorithm for separable nonlinear least-squares problems. IEEE Trans. Autom. Control **64**(2), 526–537 (2019)

94. Chen, G.Y., Gan, M., Chen, C.L.P.: Modified Gram-Schmidt method-based variable projection algorithm for separable nonlinear models. IEEE Trans. Neural Netw. Learn. Syst. **30**(8), 2410–2418 (2019)

95. Gan, M., Chen, C.L.P., Chen, G.Y., Chen, L.: On some separated algorithms for separable nonlinear squares problems. IEEE Trans. Cybern. **48**(10), 2866–2874 (2018)

96. Xiong, W., Jia, X., Yang, D., Ai, M., Li, L., Wang, S.: DP-LinkNet: A convolutional network for historical document image binarization. KSII Trans. Internet Inf. Syst. **15**(5), 1778–1797 (2021)

97. Xiong, W., Zhou, L., Yue, L., Li, L., Wang, S.: An enhanced binarization framework for degraded historical document images. EURASIP J. Image Video Process. **2021**, Article number: 13 (2021)

98. Chang, C., Wang, Q.Y., Jiang, J.C., Wu, T.Z.: Lithium-ion battery state of health estimation using the incremental capacity and wavelet neural networks with genetic algorithm. J. Energy Storag. **38**, Article Number: 102570 (2021)

99. Chang, C., Wu, Y.T., Jiang, J.C., et al.: Prognostics of the state of health for lithium-ion battery packs in energy storage applications. Energy **239**, Part B, 122189 (2022)