Automatica 137 (2022) 110117

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

A simple robust method of fractional time-delay estimation for linear dynamic systems ☆

Fengwei Chen [a], Peter C. Young [b,*]

[a] School of Automation, Chongqing University, Chongqing 400044, China
[b] Lancaster Environment Centre and Data Science Institute, Lancaster University, Lancaster LA1 4YQ, UK

ARTICLE INFO

Article history:
Received 2 October 2019
Received in revised form 12 September 2021
Accepted 22 November 2021
Available online xxxx

Keywords:
Continuous-time model
Instrumental variable
Identification
Fractional time delay

ABSTRACT

This paper is concerned with estimating the parameters of single-input, continuous-time systems from sampled input–output data, where any input time delays may be a fraction of the sampling interval. The proposed method estimates the parameters by minimizing a cost function defined by the sum of squared errors between the predicted and measured outputs using the optimal *Refined Instrumental Variable method for Continuous-time* models (RIVC). Potential difficulties may be encountered when standard gradient-based optimization is used for solving this optimization problem directly. These arise because the cost function is normally multi-modal with respect to the fractional time delay and they can occur for several reasons, such as aliasing errors, the nature of the system dynamics, and the selection of the excitation signal. In order to avoid the problems caused by these local minima, a two-stage procedure is proposed combining RIVC estimation of the model parameters with a grid search for the fractional time delay. Although this algorithm is not quite as computationally efficient as standard gradient-based optimization, it is simple, robust, and more widely applicable, as illustrated by numerical examples.

1. Introduction

Pure time delays are a common feature of many real-life systems such as chemical, environmental, biological, and industrial processes. Obtaining accurate estimates of a time delay, which may be a fraction of the sampling interval, can be of significant importance for the analysis, prediction, and control of such systems. The identification of time-delay systems has been studied for a long time and many effective methods have been available in the literature. Amongst these, there is the so-called 'implicit method', where the time delay is not an explicit parameter in the model but is approximated in some manner. Examples of this include the approximation of the time delay by a linear rational transfer function, as in Björklund (2003); and where the numerator is expanded to cover the maximum time-delay range (see, e.g., Kurz and Goedecke (1981)).

Although these implicit approaches allow for the identification of time-delay systems using conventional linear system identification methods, they have several shortcomings. For instance, they are limited to relatively small time delays and they may suffer from numerical problems in digital implementation, due to an increase in the number of parameters to be estimated. Clearly, therefore, it is more straightforward and accurate to allow for the direct estimation of fractional delays that can match the estimated pure time delay to the actual time delay with any specified level of accuracy.

The estimation of the pure time delay between two signals received at separated sensors has been considered in the control or signal processing literature, where it is based on correlation analysis, sometimes combined with conventional *Least-Squares* (LS) or subspace estimation of the rational model parameters (see, e.g., Zheng and Feng (1990)). On the other hand, one can simply estimate the rational model using standard procedures such as *Prediction-Error Minimization* (PEM) or *Refined Instrumental Variable* (RIV) estimation over a specified range of *integer* time delays and select the best model, based on the chosen identification criteria (e.g., Box and Jenkins (1970), Young (2011)). While this approach is the standard procedure in control engineering and statistical time series analysis, it clearly provides only an approximate model if the time delay is a fraction of the sampling interval.

☆ This work was partially supported by the National Natural Science Foundation of China under Grant 62073246, and by the State Key Laboratory of Intelligent Control and Decision of Complex Systems. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Cristian R. Rojas under the direction of Editor Torsten Söderström.

* Corresponding author.
E-mail addresses: fengwei.chen@cqu.edu.cn (F. Chen), p.young@lancaster.ac.uk (P.C. Young).

https://doi.org/10.1016/j.automatica.2021.110117
0005-1098/

In order to address fractional time-delay estimation, standard gradient-based optimization has already been combined with prediction- and output-error methods (Ahmed, 2020; Ljung, 2002; Maruta & Sugie, 2013); and *Instrumental Variable* (IV) methods (e.g. Baysse, Carrillo, & Habbadi, 2011; Chen, Garnier, & Gilson, 2015; Chen, Zhuan, Garnier, & Gilson, 2018; Ha, Welsh, & Alamir, 2018; Yang, Iemura, Kanae, & Wada, 2007; Young, 2014). It is well known, however, that the cost as a function of the fractional time delay is always multi-modal (Ferretti, Maffezzoni, & Scattolini, 1996; Pupeikis, 2020). This can occur for several reasons, such as the assumption made about the inter-sample behavior, the placement of the zeros and poles, and the selection of the excitation signal (Chen et al., 2018). Consequently, if poor initial parameters are selected, the parameter estimates may be trapped by local minima, making the algorithm not very robust compared with the standard integer delay estimation approach. In the present paper, we describe an approach that avoids these problems and is robust in this sense. In particular, it is insensitive to the initialization conditions specified by the user and locates the optimal estimates in a reliable manner.

The proposed method is similar to the standard integer delay identification procedure but the range of time delays is extended to include fractional delays by interpolating the input between the measured sample points and then re-sampling this extended input for a specified range of fractional time delays. As in the integer time-delay case, the best fractional time delay is then identified by repeated estimation of the model over the range of fractionally delayed inputs that are now available. The interpolation and re-sampling effectively extend the range of integer time delays to include the fractional delays at any chosen resolution.

Although this method is computationally marginally more expensive when compared to existing methods, such as that described in Chen et al. (2018), it is simple and robust in operation, as shown by the simulation examples presented in Section 6 of the paper. Before this, Section 2 formulates the identification problem; Section 3 describes the proposed method in more detail; Section 4 presents some theoretical results regarding the proposed method; and practical considerations are discussed in Section 5.

2. Problem statement

Assume that the noise free response $x_o(t)$ and input $u(t)$ of a *Single-Input, Single-Output* (SISO) system are related by the following equation:

$$x_o^{(n_a)}(t) + a_1^o x_o^{(n_a-1)}(t) + \cdots + a_{n_a}^o x_o(t) = b_0^o u^{(n_b)}(t-\tau_o) + \cdots + b_{n_b}^o u(t-\tau_o) \tag{1}$$

where the superscript or subscript 'o' designates the true value; $x_o^{(i)}(t)$ is the ith time derivative of the *Continuous-Time* (CT) signal $x_o(t)$; $\tau_o \geqslant 0$ is the pure time delay, which can be fractional; and n_a, n_b are the polynomial degrees ($n_a \geqslant n_b$). Note that $n_a = n_b$ is allowed here because the technique employed to estimate the time delay is derivative free, in contrast to the alternative approach in Chen et al. (2018). The process given in (1) can be written in a more compact transfer function form

$$x_o(t) = G_o(p)u(t-\tau_o) = \frac{B_o(p)}{A_o(p)}u(t-\tau_o) \tag{2}$$

where $B_o(p)$ and $A_o(p)$ are the system numerator and denominator polynomials defined by the differentiation operator $p : p(\cdot) = \mathrm{d}(\cdot)/\mathrm{d}t$

$$B_o(p) = b_0^o p^{n_b} + \cdots + b_{n_b}^o \tag{3a}$$

$$A_o(p) = p^{n_a} + a_1^o p^{n_a-1} + \cdots + a_{n_a}^o, \ \ n_a \geqslant n_b. \tag{3b}$$

It is further assumed that the CT input $u(t)$ is generated via a hold device from a *Discrete-Time* (DT) series $\{u(t_k)\}_{k=0}^N$, where $t_k = kT$ is the sampling instant with T the sampling interval. The measurement equation is then written as

$$y(t_k) = x_o(t_k) + \xi_o(t_k) \tag{4}$$

where $\xi_o(t_k)$ is the measurement noise that can be described by a DT *Auto-Regressive, Moving Average* (ARMA) process, i.e.,

$$\xi_o(t_k) = H_o(q^{-1})e_o(t_k) = \frac{C_o(q^{-1})}{D_o(q^{-1})}e_o(t_k) \tag{5}$$

where $e_o(t_k) \sim \mathcal{N}(0, \sigma^2)$ is a serially uncorrelated DT process with zero mean and variance σ^2 (DT white noise). The polynomials $C_o(q^{-1})$ and $D_o(q^{-1})$ are defined as follows in terms of the shift operator $q^{-1} : q^{-1}\xi_o(t_k) = \xi_o(t_{k-1})$

$$C_o(q^{-1}) = 1 + c_1^o q^{-1} + \cdots + c_{n_c}^o q^{-n_c} \tag{6a}$$

$$D_o(q^{-1}) = 1 + d_1^o q^{-1} + \cdots + d_{n_d}^o q^{-n_d}, \ \ n_d \geqslant n_c. \tag{6b}$$

This *Hybrid Box–Jenkins* (HBJ) data-generating system (Garnier & Young, 2014; Young, 2015; Young, Garnier, & Gilson, 2008) can be written in the more compact form

$$\mathcal{S} : \begin{cases} x_o(t) = G_o(p)u(t-\tau_o) \\ \xi_o(t_k) = H_o(q^{-1})e_o(t_k) \\ y(t_k) = x_o(t_k) + \xi_o(t_k). \end{cases} \tag{7}$$

Let us make the following assumptions:

Assumption 1. $G_o(p)$ is stable and compact; $H_o(q^{-1})$ is stable and invertible; the polynomial degrees n_a, n_b, n_d, and n_c are known; the system operates in open loop, with zero initial conditions.

The identification objective is to estimate the unknown plant system parameters $\{a_1^o, \ldots, a_{n_a}^o, b_0^o, \ldots, b_{n_b}^o\}$, the noise system parameters $\{d_1^o, \ldots, d_{n_d}^o, c_1^o, \ldots, c_{n_c}^o\}$, and the time delay τ_o from the sampled input–output data $\{y(t_k), u(t_k)\}_{k=0}^N$.

3. Proposed methodology

In the subsequent analysis, the model of the data-generating system is obtained by reference to the HBJ system in (7), i.e.,

$$\mathcal{M} : \begin{cases} x(t) = G(p, \boldsymbol{\theta})u(t-\tau) = \dfrac{B(p, \boldsymbol{\theta})}{A(p, \boldsymbol{\theta})}u(t-\tau) \\ \xi(t_k) = H(q^{-1}, \boldsymbol{\eta})e(t_k) = \dfrac{C(q^{-1}, \boldsymbol{\eta})}{D(q^{-1}, \boldsymbol{\eta})}e(t_k) \\ y(t_k) = x(t_k) + \xi(t_k) \end{cases} \tag{8}$$

where $\boldsymbol{\theta} = [a_1, \ldots, a_{n_a}, b_0, \ldots, b_{n_b}]^\top$ is the vector of the plant model parameters, and $\boldsymbol{\eta} = [d_1, \ldots, d_{n_d}, c_1, \ldots, c_{n_c}]^\top$ is the vector of the noise model parameters. The unknown parameters are estimated by minimizing the following prediction-error cost function:

$$\left\{\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}, \hat{\tau}\right\} = \arg\min_{\boldsymbol{\theta}, \boldsymbol{\eta}, \tau} \underbrace{\frac{1}{N+1}\sum_{k=0}^{N} \varepsilon^2(t_k)}_{V(\boldsymbol{\theta}, \boldsymbol{\eta}, \tau)} \tag{9}$$

in which $\varepsilon(t_k)$ is the prediction error at sampling instant t_k given by

$$\varepsilon(t_k) = H^{-1}(q^{-1}, \boldsymbol{\eta})\left[y(t_k) - G(p, \boldsymbol{\theta})u(t_k - \tau)\right]. \tag{10}$$

If τ is assumed to be a fixed, integer valued, time delay, the optimization problem defined in (9) can be solved by the standard *Refined IV for CT systems* (RIVC) method (Young, 2015). Moreover, if the system operates in open loop, the associated parameter

estimation error covariance matrix of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ is asymptotically block diagonal, with the off-diagonal blocks being zero (Young, 2015), so allowing for an iterative, *Pseudo-Linear Regression* (PLR) procedure for estimating $\{\boldsymbol{\theta}, \tau\}$ and $\boldsymbol{\eta}$ (see Solo (1980), Young (2015) for details).

3.1. Plant model parameter estimation

Based on the input–output data, the prediction error at the kth sampling instant (10) is expressed as (Young & Garnier, 2006)

$$\varepsilon(t_k) = y_{\mathrm{f}}^{(n_a)}(t_k) - \boldsymbol{\phi}_{\mathrm{f}}^{\top}(t_k)\boldsymbol{\theta} \tag{11}$$

where $\boldsymbol{\phi}_{\mathrm{f}}(t_k)$ is the regression vector and $(\cdot)_{\mathrm{f}}$ denotes the filtered version of $(\cdot)$, i.e.,

$$\boldsymbol{\phi}_{\mathrm{f}}^{\top}(t_k) = \left[-y_{\mathrm{f}}^{(n_a-1)}(t_k), \ldots, -y_{\mathrm{f}}(t_k), u_{\mathrm{f}}^{(n_b)}(t_k-\tau), \ldots, u_{\mathrm{f}}(t_k-\tau)\right] \tag{12}$$

with $z_{\mathrm{f}}^{(i)}(t_k) = \left[A(p,\boldsymbol{\theta})H(q^{-1},\boldsymbol{\eta})\right]^{-1} z^{(i)}(t_k)$, $z = y$ or u. The LS method is a standard procedure for estimating $\boldsymbol{\theta}$ in (11), but the resulting estimate is always biased, because $\boldsymbol{\phi}_{\mathrm{f}}(t_k)$ is correlated with the measurement noise. Here, an IV technique for CT models is adopted to address this problem, using the optimal instrument (Young, 2011)

$$\boldsymbol{\psi}_{\mathrm{f}}^{\top}(t_k) = \left[-x_{\mathrm{f}}^{(n_a-1)}(t_k), \ldots, -x_{\mathrm{f}}(t_k), u_{\mathrm{f}}^{(n_b)}(t_k-\tau), \ldots, u_{\mathrm{f}}(t_k-\tau)\right] \tag{13}$$

where $x_{\mathrm{f}}(t_k)$ denotes the filtered response of the auxiliary model parametrized by $\boldsymbol{\theta}$ and τ: $x_{\mathrm{f}}(t_k) = G(p,\boldsymbol{\theta})u_{\mathrm{f}}(t_k-\tau)$. Subsequently, the RIVC estimate of $\boldsymbol{\theta}$ is given as

$$\hat{\boldsymbol{\theta}} = \left[\sum_{k=0}^{N} \boldsymbol{\psi}_{\mathrm{f}}(t_k)\boldsymbol{\phi}_{\mathrm{f}}^{\top}(t_k)\right]^{-1} \sum_{k=0}^{N} \boldsymbol{\psi}_{\mathrm{f}}(t_k) y_{\mathrm{f}}^{(n_a)}(t_k). \tag{14}$$

This is solved in an iterative PLR manner that can be considered as a special type of Gauss–Newton optimization (Young, 2015).

3.2. Noise model parameter estimation

In the above estimation procedure, the measurement noise is estimated as

$$\hat{\xi}(t_k) = y(t_k) - G(p,\hat{\boldsymbol{\theta}})u(t_k-\tau) \tag{15}$$

where $\hat{\boldsymbol{\theta}}$ is the concurrent estimate of the CT model parameter vector. The parameters of an ARMA noise model are then estimated, based on the noise series $\{\hat{\xi}(t_k)\}$ using standard estimation routines, such as that described and used in Young (2015).

In most practical applications, the plant model is the main objective of model identification and the proposed IV-based method is able to generate statistically consistent estimates of the plant model, even if the noise model is not correctly identified (although correct identification ensures lower variance or statistically efficient estimates). Therefore, we will not investigate the properties of the estimated noise model in the next section, and just assume $H(q^{-1},\boldsymbol{\eta}) = 1$.

3.3. Summary of the algorithm

The proposed algorithm is summarized below as Algorithm 1. This is implemented using the standard rivcbj implementation of the RIVC algorithm in the CAPTAIN Toolbox for Matlab,[1] so the complete algorithm can be regarded an extension of rivcbj for fractional time-delay system identification and will be made available as the rivcbjfd routine in future versions of the Toolbox.

Algorithm 1: rivcbjfd

Input:
1. Sampled input–output data: $\{y(t_k), u(t_k)\}_{k=0}^{N}$;
2. Time-delay range: $[\tau_{\min}, \tau_{\max}]$;
3. Time-delay grid interval reduction factor: $M \in \mathbb{Z}^{+}$;
4. Anti-aliasing filter: $f(p)$, if the input is not smooth;
5. Polynomial degrees: n_a, n_b, n_d, and n_c;
6. Number of grid interval reduction: $N_{\mathrm{red}} \in \mathbb{N}$;

Output: $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\eta}}$, and $\hat{\tau}$;

filter the sampled data by $f(p)$ if the input is not smooth

$y(t_k) \leftarrow f(p)y(t_k)$, $u(t_k) \leftarrow f(p)u(t_k)$;

set the initial time delay $\hat{\tau} = 0$;

set the initial grid interval $\Delta \leftarrow T$;

for $m = 0$ **to** N_{red} **do**

  foreach $\tau_\ell = \hat{\tau} + \ell\Delta \in [\tau_{\min}, \tau_{\max}]$, $\ell \in \mathbb{Z}$, **do**

    interpolate $u(t_k - \tau_\ell)$;

    apply rivcbj to estimate $\boldsymbol{\theta}(\tau_\ell)$ [and $\boldsymbol{\eta}(\tau_\ell)$ if $m = N_{\mathrm{red}}$] from $\{y(t_k), u(t_k - \tau_\ell)\}_{k=0}^{N}$;

    compute the cost $V(\tau_\ell)$;

  end

  update the time-delay and model parameter estimates

  $\hat{\tau} = \arg\min_{\tau_\ell} V(\tau_\ell)$, $\hat{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta}(\hat{\tau})$, $\hat{\boldsymbol{\eta}} \leftarrow \boldsymbol{\eta}(\hat{\tau})$;

  shrink the time-delay range

  $[\tau_{\min}, \tau_{\max}] \leftarrow \left[\hat{\tau} - \Delta, \hat{\tau} + \Delta\right]$;

  reduce the time-delay grid interval $\Delta \leftarrow \Delta/M$;

end

Remark 1. In Algorithm 1, the RIVC method is used as the basis for the identification of models based on the interpolated and sub-sampled data, $\{y(t_k), u(t_k - \tau)\}$. Here, an anti-aliasing filter $f(p)$ is used to remove the aliasing errors in re-sampling a signal that does not have a limited band. For example, in the case of a *Zero-Order Hold* (ZOH) input signal, the fractionally delayed series $\{u(t_k - \tau)\}$ will be the same as the original if no pre-filtering is performed, so impeding the correct estimation of the fractional time delay. It should be noted that the above anti-aliasing filtering, interpolation, and sub-sampling method (referred to as the 'approximate method' from hereon) is an alternative to the exact method presented in Section 3.2 of Chen et al. (2015). This approximate method is adopted because it can save computational time compared with the exact method. However, the price paid for this improvement is that it could introduce interpolation errors which, if significant, may cause a loss in the estimation accuracy and could also complicate the theoretical analysis. Fortunately, as shown later in Section 6, the bias levels on the parameter estimates are very small. This is because of the suitable anti-aliasing filtering and interpolation procedures used in pre-filtering. Taking these considerations into account, such interpolation errors are assumed to be insignificant and so are not considered in the analysis section.

Remark 2. The first stage of Algorithm 1 is a coarse grid search on the time-delay range specified by the user, where the time-delay

[1] The CAPTAIN Toolbox for Matlab is free and can be downloaded from http://wp.lancs.ac.uk/captaintoolbox/.

resolution is set to the sampling period, i.e., $\Delta = T$, so the delayed sequence in Step 6 can be obtained by directly shifting the data. Once this first stage has been completed, the time-delay range is immediately reduced to the region between the intervals before and after the optimal time delay and, at the same time, the time-delay resolution is also divided by M to enable a fine search in the next round. In this regard, the achievable time-delay resolution is obviously $\Delta = T/M^{N_{\text{red}}}$. There is no theoretical guarantee that the above search scheme can always reach global convergence. However, it is reasonable to assume that the optimal time delay always lies in the region between the sampling intervals before and after the best integer time delay obtained from the first round; and this is confirmed if we plot the nature of the cost function, according to Theorem 1, for several typical systems. In view of this, the above scheme performs well in practice and has never failed to converge in the extensive practical evaluation we have carried out, in addition to that reported in this paper[2]. In order to ensure fast implementation of the algorithm, it is suggested that each time the rivcbj routine is called to estimate a model at the given time delay τ_ℓ, the estimated parameter vector from the previous run, i.e., $\boldsymbol{\theta}(\tau_{\ell-1})$, is used as the initial condition. Also, the maximum iteration number of rivcbj can be constrained to further reduce the computational time, if this is required.

Remark 3. The following are some guidelines regarding the choice of the hyper-parameters to initiate the algorithm. By default, one can simply let $\tau_{\min} = 0$ if no *a priori* knowledge is available; while the upper boundary can be computed based on sampled input–output data by minimizing the cross-correlation function

$$\tau_{\max} = \arg\max_{\tau} \mathbb{E}\left\{|y(t_k)u(t_k - \tau)|\right\} \tag{16}$$

where $\mathbb{E}$ is the expectation operator. Note that the resulting $\tau_{\max}$ is always larger than the true time delay due to the system phase lag. The anti-aliasing filter $f(p)$ can take various forms. As its main function is to attenuate the aliasing errors, it is intuitive that the filter bandwidth should be no larger than the Nyquist frequency. There is no general guideline for the choice of M and N_{red}, but one should be aware that the parameter estimates become more accurate when M and N_{red} increase. This is discussed further in Section 6.2 where a simulation example illustrates how M and N_{red} affect the resulting estimation of the model parameters. The choice of degrees n_a, n_b, n_d, and n_c, can be achieved by reference to statistical measures, such as the coefficient of determination R_T^2 and model order identification criteria (see Young (2011)).

4. Theoretical results

4.1. Factors influencing the cost function

In direct CT modeling from DT data, the inter-sample behavior must be assumed in order to generate output predictions and filtered derivatives. In this paper, the CT input $u(t)$ is reconstructed from a DT series $\{u(t_\ell)\}$ through B-spline interpolation

$$u(t) = \sum_{\ell=0}^{\infty} u(t_\ell)\beta_r(t - t_\ell), \;\; t_\ell = \ell T, \;\; r = 0, 1, 2, \ldots \tag{17}$$

where $\beta_r(t)$ is the rth-order B-spline defined by (see also Sánchez and Yuz (2019))

$$\beta_r(t) = \sum_{\ell=0}^{r+1} \frac{(-1)^\ell}{r!T^r} \binom{r+1}{\ell} (t - t_\ell)^r \mu(t - t_\ell) \tag{18}$$

[2] Some practical, real data examples are described in Young and Chen (2021).

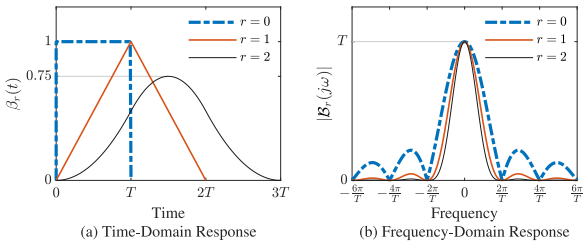

Fig. 1. Time- and frequency responses of B-splines.

and $\mu(t)$ is the unit step function. The B-spline is a generalized hold, which implements ZOH when $r = 0$, and *First-Order Hold* (FOH) when $r = 1$; see Gillberg and Ljung (2010), Sánchez and Yuz (2019) for more properties about B-splines. Moreover, $\beta_r(t)$ has the following frequency-domain description:

$$\mathcal{B}_r(j\omega) = \frac{1}{T^r}\left(\frac{1 - \mathrm{e}^{-j\omega T}}{j\omega}\right)^{r+1}. \tag{19}$$

The time- and frequency-domain responses of different B-splines, computed using (18) and (19), are shown in Fig. 1. It can be seen that the spectrum of the reconstructed CT input applied to the model is dependent on the order r. These indicate that both the estimated plant model parameters and time delay will be biased if the assumed inter-sample behavior does not match the ground-truth; see also (Pan, González, Welsh, & Rojas, 2020) for more discussions.

It is well known that the prediction- or output-error cost as a function of the fractional time delay is always multi-modal. This can occur for several reasons, such as the placement of zeros and poles (Chen et al., 2018; Ferretti et al., 1996). In this section, we investigate the impacts of inter-sample behavior, as well as the noise whitening used in the full HBJ modeling, on the cost function. From this analysis, it is clear that the proposed grid search method is a better option than the standard gradient-based search if one wants to identify the global optimum effectively.

Consider the following cost function:

$$V(\boldsymbol{\theta}, \tau) = \sum_{k=0}^{N} [y(t_k) - G(p, \boldsymbol{\theta})u(t - \tau)]^2. \tag{20}$$

Since the behavior of $V(\boldsymbol{\theta}, \tau)$ will be very complex if $\boldsymbol{\theta}$ is also considered as an unknown parameter, the following assumptions are made, for simplicity, throughout this subsection.

Assumption 2. The CT input $u(t)$ is generated by (17) from a stationary, DT process $\{u(t_k)\}$. $G_{\text{o}}(p)$ is assumed to be known, i.e., $G(p, \boldsymbol{\theta}) = G_{\text{o}}(p)$, and the output measurements are noise free.

Denote the delay-free output of $G_{\text{o}}(p)$ at time instant t_k as

$$g(t_k) = G_{\text{o}}(p)u(t_k). \tag{21}$$

Under Assumption 2, the cost function defined in (20) can be written as a function of $g(t_k)$

$$V(\tau) = \sum_{k=0}^{N} [g(t_k - \tau_{\text{o}}) - g(t_k - \tau)]^2. \tag{22}$$

It is still not straightforward to see from the above equation which factors may influence the cost function. In order to clarify this, we present a theorem which is similar to Lemma 1 in Chen, Garnier, Padilla, and Gilson (2020). Note that, unlike the case in Chen et al. (2020), we do not assume $T \to 0$, in order to take the impact of inter-sample behavior into consideration.

Theorem 1. *Under Assumption 2, let us define*

$$\bar{h}(t_k-\tau)=\int_0^{t_k} h(t_k-\tau-s)\beta_r(s)\mathrm{d}s \tag{23}$$

where $h(t)$ is the impulse response of $G_o(p)$, and $\beta_r(t)$ is the B-spline in (18); and also

$$C_{\bar{h}}(t_\ell,\tau_1,\tau_2)=\sum_{k=0}^{N}\bar{h}(t_k-\tau_1)\bar{h}(t_k-\tau_2+t_\ell) \tag{24a}$$

$$C_u(t_\ell)=\sum_{k=0}^{N}u(t_k)u(t_k+t_\ell). \tag{24b}$$

Subsequently, when the sample size N is sufficiently large, the cost function $V(\tau)$ in (22) can be written as

$$V(\tau)=\sum_{\ell=-N}^{N}\left[C_{\bar{h}}(t_\ell,\tau_o)+C_{\bar{h}}(t_\ell,\tau)-2C_{\bar{h}}(t_\ell,\tau_o,\tau)\right]C_u(t_\ell) \tag{25}$$

where, for convenience, $C_{\bar{h}}(t_\ell,\tau)=C_{\bar{h}}(t_\ell,\tau,\tau)$.

Proof. See Appendix A. □

Remark 4. In Theorem 1, $V(\tau)$ is expressed as a sum of several autocorrelation functions, from which the possible factors that may influence $V(\tau)$ are clearer: the system dynamics that determine $h(t)$; the interpolation function $\beta_r(t)$ which, together with $h(t)$, determines $\bar{h}(t_k)$; and the input signal $u(t_k)$. Unfortunately, direct computation of $V(\tau)$ in Theorem 1, using the impulse response $\bar{h}(t_k,\tau)$, is not easy, especially for high-order systems. In what follows, an alternative is presented in order to overcome this deficiency.

Proposition 1. *Assuming $n_a > n_b$, then (21) can be written in state-space form*

$$\begin{cases}\dot{\boldsymbol{x}}(t)=F_c\boldsymbol{x}(t)+G_cu(t)\\ g(t_k)=K\boldsymbol{x}(t_k)\end{cases} \tag{26}$$

where F_c, G_c, and K are the CT system matrices; $\boldsymbol{x}(t)$ is the state vector and $\dot{\boldsymbol{x}}(t)$ is its time derivative. As in (23) and (24a), let us define

$$\bar{\boldsymbol{h}}(t_k-\tau)=\int_0^{t_k}\boldsymbol{h}(t_k-\tau-s)\beta_r(s)\mathrm{d}s \tag{27a}$$

$$C_{\bar{\boldsymbol{h}}}(\tau_1,\tau_2)=\sum_{k=0}^{N}\bar{\boldsymbol{h}}(t_k-\tau_1)\bar{\boldsymbol{h}}^{\top}(t_k-\tau_2) \tag{27b}$$

where $\boldsymbol{h}(t)=\mathrm{e}^{F_ct}G_c\mu(t)$, with $\mu(t)$ the unit step, is the impulse response of the state equation (26), and $\beta_r(t)$ is given in (18). Under the ZOH interpolation and $N\to\infty$, it can be shown that

$$C_{\bar{\boldsymbol{h}}}(\tau_1,\tau_2)=\begin{cases}F_d(\bar{\tau}_1-\tau_1)PF_d^{\top}(\bar{\tau}_1-\tau_2)\\ \quad+G_d(\bar{\tau}_1-\tau_1)G_d^{\top}(T)F_d^{\top}(\bar{\tau}_1-\tau_2-T),\ \bar{\tau}_1>\bar{\tau}_2\\ F_d(\bar{\tau}_1-\tau_1)PF_d^{\top}(\bar{\tau}_1-\tau_2)\\ \quad+G_d(\bar{\tau}_1-\tau_1)G_d^{\top}(\bar{\tau}_1-\tau_2),\ \bar{\tau}_1=\bar{\tau}_2\\ C_{\bar{\boldsymbol{h}}}^{\top}(\tau_2,\tau_1),\ \bar{\tau}_1<\bar{\tau}_2\end{cases} \tag{28}$$

where $\bar{\tau}$ is the smallest integer time delay larger than or equal to τ; $F_d(t)=\mathrm{e}^{F_ct}$; $G_d(t)=\int_0^t\mathrm{e}^{F_cs}\mathrm{d}s\,G_c$; and P is the solution to the DT Lyapunov equation $P=F_d(T)PF_d^{\top}(T)+G_d(T)G_d^{\top}(T)$.

Proof. See Appendix B. □

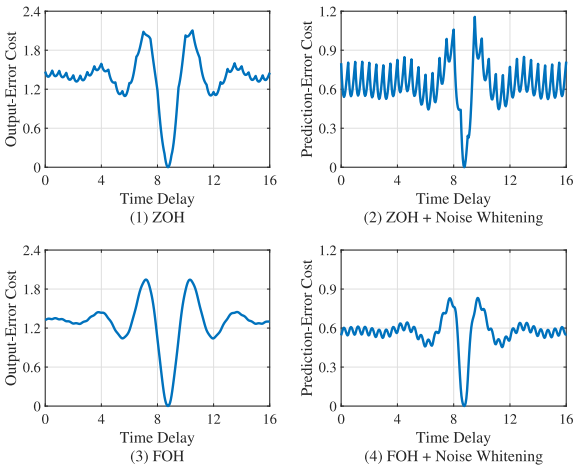

Fig. 2. Plots of the output- and prediction-error cost functions.

By Proposition 1, the autocorrelations in (25) are computed as

$$C_{\bar{h}}(t_\ell,\tau_1,\tau_2)=KC_{\bar{\boldsymbol{h}}}(\tau_1,\tau_2-t_\ell)K^{\top}. \tag{29}$$

Remark 5. For FOH interpolation, it should be noted that $u'(t_k)=[u(t_{k+1})-u(t_k)]/T$ is ZOH and, correspondingly, $C_{\bar{\boldsymbol{h}}}(\tau_1,\tau_2)$ can be computed based on the increased order transfer function $G_o(p)/p$ and the ZOH input series $u'(t_k)$, via the same procedure as presented in Proposition 1. For high-order B-spline interpolation, the underlying ZOH input series can be reconstructed in a similar but slightly more complicated manner, see Corollary 3.5 of Sánchez and Yuz (2019).

Example 1. As an illustration of the above results, let us consider the following system:

$$y(t)=G_o(p)u(t-\tau_o)=\frac{p+3}{p^2+p+4}u(t-8.75) \tag{30}$$

where the CT input is generated by interpolating a DT white noise process of unit variance with the ZOH or FOH assumption. The sampling interval is chosen as $T=0.5$. When FOH is assumed, the idea in Remark 5 is used to reconstruct the underlying ZOH input series, i.e., $u'(t_k)=[u(t_{k+1})-u(t_k)]/T$. It can be shown that FOH is equivalent to the following modified B-spline: $\mathrm{e}^{j\omega T}\mathcal{B}_1(j\omega)$, where $\mathcal{B}_1(j\omega)$ is defined in (19). Moreover, if a noise model is used to compute the prediction-error cost, e.g., $H(q^{-1},\boldsymbol{\eta})=1/(1-1.228q^{-1}+0.6065q^{-2})$, the input series should also be filtered through the inverse noise model $H^{-1}(q^{-1},\boldsymbol{\eta})$. Subsequently, the output- and prediction-error costs are computed according to Theorem 1 and Proposition 1, as shown in Fig. 2, where numerous local minima are observed when ZOH is assumed, each one belonging to an interval of T; while FOH is able to smooth the cost to some satisfactory extent. However, using the inverse noise model to whiten the noisy data severely deteriorates these cost functions. This is reasonable, since the inverse noise model is, by nature, a high-pass filter, which cancels the low-pass filtering effect provided by the plant and thus magnifies the aliasing effect. As a consequence, it seems that a grid search of the kind we propose could be the most effective and robust method for identifying the optimal fractional time delay in this situation.

4.2. Persistence of excitation

In the application of the proposed method, the system should be persistently excited in order to generate informative data for

model identification. This ensures that the matrix to be inverted in (14) is well conditioned and its inversion exists. Previous research (Chen et al., 2018) has investigated the persistent excitation conditions under multi-sine input assumptions, where the problem was simplified to determine a necessary amount of constituent harmonics that would ensure model identifiability. The analysis therein imposed the sampling period $T \to 0$, but it can be shown that the results still hold in the case where the sampling period is not zero.

Interesting research on the identifiability of CT models from sampled data is presented in Pan et al. (2020), where the DT equivalent model is used as a medium to aid the identifiability analysis. However, the DT model can be considered as the hold operation plus the CT model and this has an increased number of parameters, so requiring a higher-level of persistent excitation in the analysis of CT model identifiability. In this paper, therefore, the persistence of excitation for direct CT modeling is investigated in an alternative manner, without reference to the equivalent DT model. Before this, the following assumptions are made.

Assumption 3. The DT input series $\{u(t_k)\}$ is sampled instantaneously from a multi-sine $u_0(t) = \sum_{\ell=1}^{L} \alpha_\ell \sin(\omega_\ell t + \gamma_\ell)$ with sampling interval T, where $0 < \omega_\ell < \pi/T$ ($\omega_\ell \neq \omega_n$ if $\ell \neq n$); $\alpha_\ell \neq 0$ and γ_ℓ are real numbers. It is further assumed that $H(q^{-1}, \boldsymbol{\eta}) = 1$, and $\{u(t_k)\}$ is uncorrelated with the measurement noise $\{\xi(t_k)\}$.

Assumption 4. The CT input–output signals are reconstructed from DT data via (17) using the assumptions about inter-sample behaviors of the input–output signals. The interpolation errors are negligible when the model identification is based on the fractionally delayed series $\{y(t_k), u(t_k - \tau)\}$.

The following theorem follows when these assumptions are satisfied, where the asymmetric, normal matrix converges to a symmetrical matrix at the true parameter values $\{\boldsymbol{\theta}_o, \tau_o\}$, the non-singularity of which is easier to prove. Note that the rivcbj algorithm uses the symmetric version of the RIVC algorithm in a final iteration, after convergence, as required by the Pierce theorem (Pierce, 1972; Young, 2015). Moreover, such a strict assumption is only for brevity of theoretical analysis, and experience shows that, up to convergence, the asymmetric matrix is always non-singular for a wide range of $\{\boldsymbol{\theta}, \tau\}$ values.

Theorem 2. *Under Assumption 1, 3, and 4, when* $\{\boldsymbol{\theta}, \tau\} = \{\boldsymbol{\theta}_o, \tau_o\}$, *the normal matrix* $\lim_{N\to\infty} \frac{1}{N+1} \sum_{k=0}^{N} \boldsymbol{\psi}_f(t_k) \boldsymbol{\phi}_f^\top(t_k)$ *is non-singular if the number of constituent harmonics in* $u_0(t)$ *satisfies* $L \geqslant (n_a + n_b + 1)/2$.

Proof. See Appendix C. □

4.3. Accuracy

This subsection provides a brief analysis of the final convergence point. Let us consider the alternate formulation of the RIVC optimization problem:

$$\left\{\hat{\boldsymbol{\theta}}, \hat{\tau}\right\} = \arg\min_{\boldsymbol{\theta},\tau} \left\| \mathbb{E}\left\{\boldsymbol{\psi}_f(t_k)\left[y(t_k) - G(p)u(t_k - \tau)\right]\right\}\right\|^2. \tag{31}$$

Since the measurement noise is zero mean and uncorrelated to $\boldsymbol{\psi}_f(t_k)$, it is straightforward to see that the global optimum of problem (31) is the solution to the equation: $G_o(p)u(t_k - \tau_o) - G(p)u(t_k - \tau) = 0$. This equation can be written in the frequency domain as

$$\left[A(j\omega)B_o(j\omega) - A_o(j\omega)B(j\omega)e^{j\omega(\tau_o - \tau)}\right] \frac{U(j\omega)e^{-j\omega\tau_o}}{A_o(j\omega)A(j\omega)} = 0. \tag{32}$$

Let $\mathcal{D} = \{\ell T/M : \ell \in \mathbb{N}, M \in \mathbb{Z}^+\}$ be a set of candidate values for τ. It is assumed that the conditions in Theorem 2 are fulfilled. On one hand, if $\tau_o \in \mathcal{D}$, there exist $\tau \in \mathcal{D}$ so that $\tau - \tau_o = 0$ and then we need to solve $A(j\omega)B_o(j\omega) - A_o(j\omega)B(j\omega) = 0$. As $G_o(j\omega)$ and $G(j\omega)$ are compact, i.e., $A(j\omega)$ and $B(j\omega)$ [$A_o(j\omega)$ and $B_o(j\omega)$] do not have any common factor, and the degrees n_a and n_b are known, it can be proven that $\boldsymbol{\theta}_o$ is the unique solution.

On the other hand, if $\tau_o \notin \mathcal{D}$ then $\tau - \tau_o \neq 0$, the following polynomial is of infinite order:

$$A(j\omega)B_o(j\omega) - A_o(j\omega)B(j\omega)e^{-j\omega(\tau - \tau_o)} \tag{33}$$

because $e^{j\omega(\tau_o - \tau)} = \sum_{\ell=0}^{\infty} [j\omega(\tau_o - \tau)]^\ell / \ell!$, so we cannot find a $\hat{\boldsymbol{\theta}}$ to make (33) zero. As a result, the estimated $\hat{\boldsymbol{\theta}}$ is biased. However, we have $e^{-j\omega(\tau - \tau_o)} \to 1$ if $\tau - \tau_o \to 0$, indicating that $\hat{\boldsymbol{\theta}}$ will converge to $\boldsymbol{\theta}_o$ at this limit. Therefore, it is reasonable to select a large M, which reduces the time-delay resolution $\Delta = T/M$, so as to improve the accuracy of $\hat{\boldsymbol{\theta}}$.

5. Practical considerations

There are various ways in which the proposed method can be applied in practice. The examples considered in Section 6 make use of the rivcbjid and rivcbjfd routines available in the CAPTAIN Toolbox (see Section 3), as well as the interp1 routine in Matlab. In all the examples, the rivcbjid structure identification routine is used to identify the best integer time-delay model from a user-specified range of model structures with different parameterizations and integer time delays.

Based on the above information, the rivcbjfd routine is then applied to see if the introduction of a fractional time delay improves the explanation of the data and, if it is considered important, that the mechanistic interpretation of the model is acceptable. Two user-defined inputs are required by rivcbjfd:

1. The interp1 routine in Matlab is exploited by rivcbjfd for the interpolation and sub-sampling of the input series. This requires the specification of an interpolation method. Our experience suggests that the 'linear', 'spline', and 'pchip' options provide the best performance.
2. The accuracy required for the identification, as defined by the number of sub-samples specified over each sampling interval.

As discussed previously, the input signal should be reasonably smooth. If this is clearly not the case (e.g., a signal with numerous step changes and ZOH inter-sample behavior), then prior low-pass filtering may need to be applied to the input and output signals. While reasonable results may be obtained without such filtering, it is safer to apply it in order to ensure satisfactory estimation (see Section 6). Finally, it is often worth using the returned information from rivcbjfd to plot the nature of the output-error variance cost as a function of the fractional time delay to see how well this is identified.

6. Simulation example

Consider the following system:

$$\begin{cases} x_o(t) = G_o(p)u(t - \tau_o) = \dfrac{-4p + 1}{p^2 + p + 4} u(t - 8.7) \\ y(t_k) = x_o(t_k) + \xi_o(t_k). \end{cases} \tag{34}$$

The input signal is a *Pseudo-Random Binary Sequence* (PRBS) of maximal length, with the number of shift registers $N_s = 11$, and FOH is used to reconstruct the CT input. The sampling interval and number of observations are 0.2 and 2047, respectively. $\xi_o(t_k)$ is white and Gaussian, and its variance is adjusted to achieve a

Table 1
Parameter estimates computed from 500 MC runs.

Method	Time-delay resolution	Estimated parameter value					R_T^2	T_c[***]	P_{gc}
		a_1 (1)	a_2 (4)	b_0 (−4)	b_1 (1)	τ (8.7)			
rivcbjfd	0.0002 ($N_{red} = 3, M = 10$)	1.0005	4.0013	−4.0013	1.0020	8.7003	0.9699	0.6912 s	100%
Standard Error		±0.0124	±0.0209	±0.0390	±0.0683	±0.0062	±0.0003		
procest[*,**]	–	1.0006	4.0000	−3.9984	1.0013	8.7000	0.9699	0.5530 s	28.8%
Standard Error		±0.0124	±0.0195	±0.0383	±0.0709	±0.0064	±0.0003		

[*]The means and standard deviations are computed from the globally convergent models.
[**]The initial delays are drawn from a uniform distribution on the interval [5, 9].
[***]T_c is the mean computational time for each MC run.

Signal-to-Noise Ratio (SNR) of 15 dB. A *Monte–Carlo* (MC) simulation of 500 runs is conducted, each run having a different realization for $\{\xi_0(t_k)\}$. The model parameters are estimated by the two schemes: the rivcbjfd routine from the CAPTAIN Toolbox; and the procest routine from the Matlab System Identification Toolbox.

The coefficient of determination R_T^2, defined as follows:

$$R_T^2 = 1 - \frac{\|y(t_k) - x(t_k)\|_2^2}{\|y(t_k) - \text{mean}\{y(t_k)\}\|_2^2} \tag{35}$$

is used as the performance index. In order to allow for uniformity in the examples, $x(t_k)$ is always computed using the Matlab routine compare. The convergence performance is assessed using P_{gc}, where $P_{gc} = N_{gc}/N_{mc} \cdot 100\%$, in which N_{gc} is the number of globally convergent models and N_{mc} is the total number of MC runs. In this example, a model is said to be globally convergent if the estimated SNR is 15 dB, which corresponds to an $R_T^2 \geqslant 0.96$, with the R_T^2 defined in (35).

As the input in this example is a non-smooth PRBS, the sampled input–output data are filtered by a 5th-order Butterworth filter of cutoff frequency 2.4 rad/s before they are used for model identification. The filter bandwidth is small compared with the system bandwidth (22.75 rad/s), but this is quite reasonable because majority of the system output power is in the low-frequency range of [0, 3] rad/s. The interp1 routine with the 'spline' option is then exploited in the rivcbjfd algorithm to compute the fractionally delayed input series $\{u(t_k - \tau)\}$ from the pre-filtered data.

6.1. Robustness to initialization

A major advantage of the proposed rivcbjfd algorithm is its robustness to the initial specification of the time delay by the user. Gradient-based methods such as procest are sensitive to this initialization because, as we have seen in previous sections, they can be affected badly by local minima. Fig. 3 illustrates this problem: it shows the results of MC analysis involving 500 realizations where the initial delay, for each MC realization, is a random number drawn from a uniform distribution on the interval [5, 9]. Note that this does not apply to rivcbjfd because it only requires the specification of a time-delay range, not a precise value. The estimated time delays are shown in the left panels and the resulting step responses are plotted in the right panels.

It is clear from these MC results that procest is very sensitive to the initial condition selection, with the resulting $P_{gc} = 28.8\%$. In contrast to this, the rivcbjfd algorithm has no failures with $P_{gc} = 100\%$, where the hyper-parameters are chosen as $M = 3$, $N_{red} = 10$, $\tau_{min} = 0$ and $\tau_{max} = 10$ (see Remark 3). The accurate estimation of the true time delay of 8.7 is possible because the resolution of the time delay in the grid search is 0.0002 and so rivcbjfd is able to locate it straightforwardly for all the realizations. The MC identification results are summarized

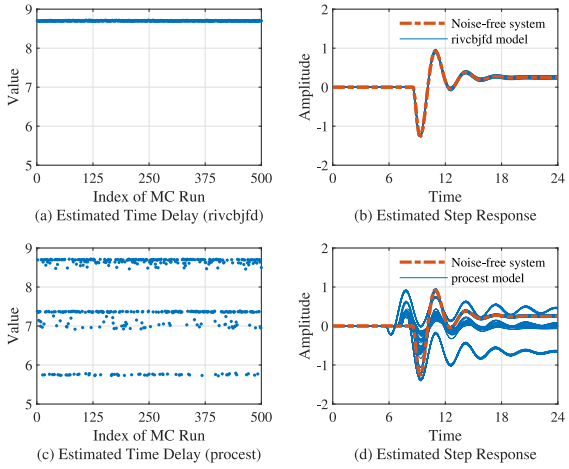

Fig. 3. MC simulation results generated by rivcbjfd and procest, where $\tau_o = 8.7$ and, for the latter method, initial delays are drawn from a uniform distribution on the interval [5, 9].

in Table 1, which shows the means and standard deviations of the parameter estimates. It is important to note that the results presented here for the procest routine are based solely on the 144 *successful* realizations and, as would be expected, these are virtually the same as the rivcbjfd parameter estimates based on all 500 realizations.

6.2. Specifying the resolution of the grid search

Although the proposed rivcbjfd algorithm requires the user to specify the resolution of the grid search, the algorithm is not too sensitive in this regard. It produces sensible results with relatively low resolutions and a fine grid is only required if particularly accurate results are required. An illustration of this is given in Table 2, which presents the MC identification results, based on 500 realizations, for the identification of the system (34), where the time-delay resolution is taken sequentially from the set {0.0667, 0.0222, 0.0074}, which corresponds to the use of the following hyper-parameter settings: $N_{red} = \{1, 2, 3\}$ and $M = 3$.

It is clear from these results that, even at the lowest resolution, the results are reasonable but that increased resolutions refine the estimates and, not surprisingly given the simulated time delay of $\tau = 8.7$, a resolution of 0.0074 or less is required to match the results obtained by gradient-based optimization using the procest routine. However, the computational time for such a search on a standard computer is still only 0.4718 s at the resolution 0.0074 (marginally larger than procest), so this does not present any problems for the user.

Table 2
Parameter estimates computed from 500 MC runs.

Method	Time-delay resolution	Estimated parameter value					R_T^2	T_c	P_{gc}
		a_1 (1)	a_2 (4)	b_0 (−4)	b_1 (1)	τ (8.7)			
rivcbjfd	0.0667 ($N_{red} = 1, M = 3$)	1.0002	4.0043	−4.0018	1.0121	8.7015	0.9654	0.4116 s	100%
Standard Error		±0.0378	±0.0833	±0.1096	±0.2438	±0.0333	±0.0006		
rivcbjfd	0.0222 ($N_{red} = 2, M = 3$)	0.9994	3.9994	−3.9993	1.0001	8.7002	0.9696	0.4457 s	100%
Standard Error		±0.0156	±0.0300	±0.0476	±0.0937	±0.0111	±0.0003		
rivcbjfd	0.0074 ($N_{red} = 3, M = 3$)	0.9996	3.9995	−3.9996	0.9999	8.7001	0.9699	0.4718 s	100%
Standard Error		±0.0127	±0.0204	±0.0403	±0.0686	±0.0064	±0.0003		
procest*	–	0.9997	3.9991	−3.9992	0.9993	8.6999	0.9699	0.2658 s	100%
Standard Error		±0.0124	±0.0196	±0.0399	±0.0667	±0.0061	±0.0003		

*The initial delay for procest in each run is fixed as 8.5.

7. Conclusions

In this paper, we have addressed the problem of parameter and fractional time-delay estimation for a SISO CT transfer function model from sampled input–output data. In order to understand the potential difficulties of estimating fractional time delays, the properties of the cost function have been investigated and it has been shown that the assumption about the inter-sample behavior, the nature of the system dynamics, and the selection of the input signal may create local minima in the cost function. As a result, the gradient-based search, a popular technique to handle fractional time-delay estimation, may fail if the initial parameters are not selected suitably. In order to relax the dependence on the initial parameter estimates, we have exploited a simple and robust approach that is similar to the standard identification procedure for integer delay systems in which the time delay is estimated via a grid search, but where the range of time delays is extended to include fractional delays by interpolating the inputs between the measured sample points and then re-sampling these extended inputs for a specified range of fractional time delays. Since the proposed estimation procedure requires the interpolation of the input signal, the nature of the assumed inter-sample behavior may, in certain circumstances, prevent the estimation of an 'alias-free' fractional time delay. However, this problem only occurs when the input signal includes sharp changes, as in step and PRBS input signals, and the paper has shown that it can be circumvented quite simply by the use of a low-pass (anti-aliasing) filter in the interpolation and identification process.

Acknowledgments

We wish to thank anonymous reviewers whose helpful comments have improved the paper.

Appendix A. Proof of Theorem 1

Following (17) and (21), the delayed signal $g(t_k - \tau)$ is computed as ($u(t) = 0$ when $t < 0$, as imposed in Assumption 1)

$$\begin{aligned} g(t_k - \tau) &= \int_0^{t_k} h(t_k - \tau - s)u(s)\mathrm{d}s \\ &= \sum_{\ell=0}^{k} \int_0^{t_k} h(t_k - \tau - s)\beta_r(s - t_\ell)\mathrm{d}s\; u(t_\ell) \\ &= \sum_{\ell=0}^{k} \overline{h}(t_k - \tau - t_\ell)u(t_\ell) = \overline{h}(t_k - \tau) * u(t_k) \end{aligned} \tag{A.1}$$

in which $*$ is the convolution operator. Since $\{u(t_k)\}$ is a stationary process, as imposed in Assumption 2, and the system is linear time invariant, it is obvious that the series $\{g(t_k - \tau)\}$ is also stationary when τ is fixed. Then, according to the relationship between the convolution and autocorrelation functions, we have

$$\begin{aligned} C_g(t_\ell, \tau_1, \tau_2) &= \sum_{k=0}^{N} g(t_k - \tau_1)g(t_k - \tau_2 + t_\ell) \\ &= g(t_\ell - \tau_1) * g(-t_\ell + \tau_2) \\ &= \overline{h}(t_\ell - \tau_1) * u(t_\ell) * \overline{h}(-t_\ell + \tau_2) * u(-t_\ell) \\ &= C_{\overline{h}}(t_\ell, \tau_1, \tau_2) * C_u(t_\ell). \end{aligned} \tag{A.2}$$

With the above relation in mind and when N is sufficiently large, $V(\tau)$ defined in (22) can be written alternatively as

$$\begin{aligned} V(\tau) &= C_g(0, \tau_o) + C_g(0, \tau) - 2C_g(0, \tau_o, \tau) \\ &= \sum_{\ell=-N}^{N} \left[C_{\overline{h}}(t_\ell, \tau_o) + C_{\overline{h}}(t_\ell, \tau) - 2C_{\overline{h}}(t_\ell, \tau_o, \tau)\right] C_u(0 - t_\ell) \\ &= \sum_{\ell=-N}^{N} \left[C_{\overline{h}}(t_\ell, \tau_o) + C_{\overline{h}}(t_\ell, \tau) - 2C_{\overline{h}}(t_\ell, \tau_o, \tau)\right] C_u(t_\ell) \end{aligned} \tag{A.3}$$

where $C_g(t_\ell, \tau) = C_g(t_\ell, \tau, \tau)$ and, in the derivation of the last equality, we have used the property: $C_u(t_\ell) = C_u(-t_\ell)$. □

Appendix B. Proof of Proposition 1

Under the ZOH assumption, the zero-order B-spline $\beta_0(t)$ is adopted for interpolation, i.e., $\beta_0(t) = 1$ if $t \in [0, T)$, otherwise 0. Then, $\overline{\boldsymbol{h}}(t_k - \tau)$ is solved as

$$\begin{aligned} \overline{\boldsymbol{h}}(t_k - \tau) &= \int_0^T \mathrm{e}^{F_c(t_k - \tau - s)}\mu(t_k - \tau - s)\mathrm{d}s\; G_c \\ &= \begin{cases} \mathbf{0}, & t_k - \tau < 0 \\ \displaystyle\int_0^{t_k - \tau} \mathrm{e}^{F_c s}\mathrm{d}s\; G_c, & 0 \leqslant t_k - \tau < T \\ \mathrm{e}^{F_c(t_k - \tau - T)} \displaystyle\int_0^T \mathrm{e}^{F_c s}\mathrm{d}s\; G_c, & t_k - \tau \geqslant T. \end{cases} \end{aligned} \tag{B.1}$$

Letting $t_M = \max\{\overline{\tau}_1, \overline{\tau}_2\}$, we have

$$\begin{aligned} C_{\overline{\boldsymbol{h}}}(\tau_1, \tau_2) &= \sum_{k=0}^{N} \overline{\boldsymbol{h}}(t_k - \tau_1)\overline{\boldsymbol{h}}^\top(t_k - \tau_2) \\ &= \underbrace{\overline{\boldsymbol{h}}(t_M - \tau_1)\overline{\boldsymbol{h}}^\top(t_M - \tau_2)}_{C_1} + \underbrace{\sum_{k=M+1}^{N} \overline{\boldsymbol{h}}(t_k - \tau_1)\overline{\boldsymbol{h}}^\top(t_k - \tau_2)}_{C_2}. \end{aligned} \tag{B.2}$$

In view of (B.1), C_1 on the right-hand side of (B.2) is readily computed as (here we first consider the case when $\overline{\tau}_1 \geqslant \overline{\tau}_2$)

$$C_1 = \overline{\boldsymbol{h}}(\overline{\tau}_1 - \tau_1)\overline{\boldsymbol{h}}^\top(\overline{\tau}_1 - \tau_2) \tag{B.3}$$

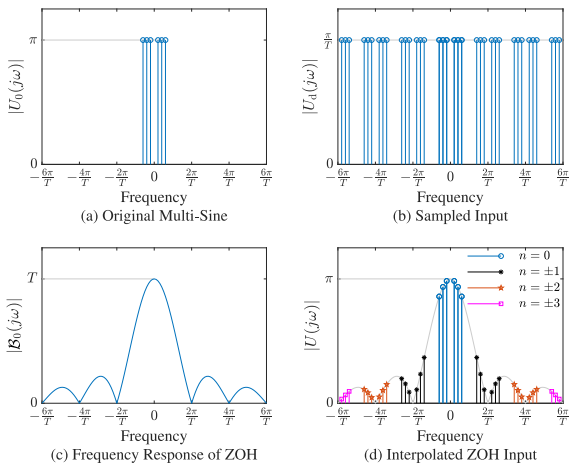

Fig. C.4. Spectrum before and after sampling and ZOH interpolation. $U_0(j\omega)$, $U_d(j\omega)$, and $U(j\omega)$ are the Fourier transforms of the multi-sine $u_0(t)$ defined in Assumption 3 (here $\alpha_\ell = 1$ for $\ell = 1, 2, 3$), the sampled input $\sum_{k=0}^{\infty} u(t_k)\delta(t - t_k)$, and the interpolated ZOH input $u(t)$ defined in (C.1), respectively.

$$= \begin{cases} G_d(\overline{\tau}_1 - \tau_1)G_d^\top(T)F_d^\top(\overline{\tau}_1 - \tau_2 - T), & \overline{\tau}_1 > \overline{\tau}_2 \\ G_d(\overline{\tau}_1 - \tau_1)G_d^\top(\overline{\tau}_1 - \tau_2), & \overline{\tau}_1 = \overline{\tau}_2. \end{cases} \tag{B.4}$$

The second term C_2 is slightly more complicated to compute. Following (B.1) and when $N \to \infty$, we have

$$\begin{aligned} C_2 &= \sum_{k=M+1}^{\infty} F_d(t_k - \tau_1 - T)G_d(T)G_d^\top(T)F_d^\top(t_k - \tau_2 - T) \\ &= F_d(\overline{\tau}_1 - \tau_1)\underbrace{\sum_{k=0}^{\infty} F_d(t_k)G_d(T)G_d^\top(T)F_d^\top(t_k)}_{P}F_d^\top(\overline{\tau}_1 - \tau_2) \\ &= F_d(\overline{\tau}_1 - \tau_1)PF_d^\top(\overline{\tau}_1 - \tau_2). \end{aligned} \tag{B.5}$$

Note that P can be decomposed by iteration as follows:

$$\begin{aligned} P &= \sum_{k=0}^{\infty} F_d(t_k)G_d(T)G_d^\top(T)F_d^\top(t_k) \\ &= \sum_{k=1}^{\infty} F_d(t_k)G_d(T)G_d^\top(T)F_d^\top(t_k) + G_d(T)G_d^\top(T) \\ &= F_d(T)PF_d^\top(T) + G_d(T)G_d^\top(T) \end{aligned} \tag{B.6}$$

which is, in fact, a DT Lyapunov equation that can be solved by the Matlab routine dlyap. The case when $\overline{\tau}_1 < \overline{\tau}_2$ can be handled following the same procedure, since we have $C_{\overline{h}}(\tau_1, \tau_2) = C_{\overline{h}}^\top(\tau_2, \tau_1)$ by definition. Finally, the substitution of (B.3) and (B.5) into (B.2) yields (28). □

Appendix C. Proof of Theorem 2

Following Assumptions 3 and 4, the CT input $u(t)$ applied to the system has the following time-domain expression (see also Fig. C.4 for an illustration of the interpolation process):

$$u(t) = \sum_{\ell=1}^{L}\sum_{n=0}^{\infty} \alpha_{\ell,n}\sin\left(\omega_{\ell,n}t + \gamma_{\ell,n}\right) \tag{C.1}$$

where $\alpha_{\ell,n}$ and $\gamma_{\ell,n}$ are the amplitude and phase of each frequency $\omega_{\ell,n} = \omega_\ell + 2\pi n/T$. The harmonics where $n \neq 0$ are created by the hold operation. The autocorrelation and power spectral density functions of the reconstructed CT input $u(t)$, denoted by $R_u(s)$ and $\Phi_u(\omega)$, respectively, are defined as

$$R_u(s) = \mathbb{E}\{u(t)u(t+s)\} = \sum_{\ell=1}^{L}\sum_{n=0}^{\infty} \frac{\alpha_{\ell,n}^2}{2}\cos\left(\omega_{\ell,n}s\right) \tag{C.2a}$$

$$\Phi_u(\omega) = \int_{-\infty}^{\infty} R_u(s)e^{-j\omega s}ds = \pi\sum_{\ell=-L}^{L}\sum_{n=-\infty}^{\infty} \frac{\alpha_{\ell,n}^2}{2}\delta\left(\omega - \omega_{\ell,n}\right). \tag{C.2b}$$

In (C.2b), when ℓ is negative, e.g., $\ell = -L$, we have $\omega_{-L,n} = -\omega_L + 2\pi n/T$. Moreover, since $\Phi_u(\omega)$ is real valued and symmetric, we have $\alpha_{\ell,-n} = \alpha_{-\ell,n}$. Similarly, the autocorrelation and power spectral density functions of the CT instrument $\boldsymbol{\psi}_f(t)$ are defined as

$$R_{\boldsymbol{\psi}}(s) = \mathbb{E}\left\{\boldsymbol{\psi}_f(t)\boldsymbol{\psi}_f^\top(t+s)\right\}, \quad \Phi_{\boldsymbol{\psi}}(\omega) = \int_{-\infty}^{\infty} R_{\boldsymbol{\psi}}(s)e^{-j\omega s}ds. \tag{C.3}$$

Moreover, $\boldsymbol{\psi}_f(t)$ can be written as

$$\begin{aligned} \boldsymbol{\psi}_f(t) &= \left[-B(p)[p^{n_a-1}, \ldots, 1], A(p)[p^{n_b}, \ldots, 1]\right]^\top \frac{1}{A^2(p)}u(t-\tau) \\ &= S\left[p^{n_a+n_b}, \ldots, 1\right]^\top \frac{1}{A^2(p)}u(t-\tau) \end{aligned} \tag{C.4}$$

where $X(p) = X(p, \boldsymbol{\theta})$, $X = A$ or B, for notational simplicity; and

$$S = \begin{bmatrix} 0 & -b_0 & \cdots & -b_{n_b} & & 0 \\ & & \ddots & & \ddots & \\ 0 & & & -b_0 & \cdots & -b_{n_b} \\ 1 & a_1 & \cdots & a_{n_a} & & 0 \\ & \ddots & \ddots & & \ddots & \\ 0 & & 1 & a_1 & \cdots & a_{n_a} \end{bmatrix}. \tag{C.5}$$

Based on the above equations and according to Ljung (1999), the expression for $\Phi_{\boldsymbol{\psi}}(\omega)$ is obtained straightforwardly in the following form:

$$\Phi_{\boldsymbol{\psi}}(\omega) = \frac{\Phi_u(\omega)}{|A^2(j\omega)|^2}\cdot S\begin{bmatrix}(j\omega)^{n_a+n_b}\\ \vdots \\ 1\end{bmatrix}\begin{bmatrix}(j\omega)^{n_a+n_b}\\ \vdots \\ 1\end{bmatrix}^{\mathrm{H}} S^\top \tag{C.6}$$

where $(\cdot)^{\mathrm{H}}$ denotes the complex conjugate transpose of $(\cdot)$. Before proceeding, some important properties of $\Phi_{\boldsymbol{\psi}}(\omega)$ should be mentioned, i.e., the following matrices are symmetrical, and positive definite if $L \geqslant (n_a + n_b + 1)/2$, with non-negative diagonal elements:

$$\frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} \Phi_{\boldsymbol{\psi}}\left(\omega + \frac{2\pi n}{T}\right)d\omega, \quad n = 0 \tag{C.7a}$$

$$\frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} \Phi_{\boldsymbol{\psi}}\left(\omega + \frac{2\pi n}{T}\right) + \Phi_{\boldsymbol{\psi}}\left(\omega - \frac{2\pi n}{T}\right)d\omega, \quad n \in \mathbb{N}^+. \tag{C.7b}$$

It is necessary, however, to prove that these matrices are also non-singular. Due to the page limits, only the proof for the case $n = 0$ is presented here, but the proofs for the other cases are quite similar. Following (C.6) and (C.2b), (C.7a) can further be written as

$$\begin{aligned} \frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} \Phi_{\boldsymbol{\psi}}(\omega)d\omega &= \sum_{\ell=-L}^{L} \frac{\alpha_{\ell,0}^2}{4|A^2(j\omega_{\ell,0})|^2} \\ &\times S\begin{bmatrix}(j\omega_{\ell,0})^{n_a+n_b}\\ \vdots \\ 1\end{bmatrix}\begin{bmatrix}(j\omega_{\ell,0})^{n_a+n_b}\\ \vdots \\ 1\end{bmatrix}^{\mathrm{H}} S^\top = S\Upsilon W\Upsilon^{\mathrm{H}}S^\top \end{aligned} \tag{C.8}$$

where

$$\Upsilon = \begin{bmatrix} (j\omega_{-L,0})^{n_a+n_b} & \cdots & (j\omega_{L,0})^{n_a+n_b} \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \tag{C.9a}$$

$$W = \frac{1}{4}\mathrm{diag}\left(\frac{\alpha^2_{-L,0}}{|A^2(j\omega_{-L,0})|^2}, \ldots, \frac{\alpha^2_{L,0}}{|A^2(j\omega_{L,0})|^2} \right). \tag{C.9b}$$

Note that S has shown to be non-singular if $B(p)$ and $A(p)$ are co-prime (see Söderström and Stoica (1983)). W is a diagonal matrix of full rank. Υ is a Vandermonde matrix which has full row rank if the number of columns is equal to, or larger than, the number of rows, i.e., $2L \geqslant n_a+n_b+1$. These together show that the matrix in (C.8) is non-singular.

In the digital implementation of the estimation algorithm, we have only the sampled instruments $\{\boldsymbol{\psi}_{\mathrm{f}}(t_k)\}$, so the discrete-time Fourier transform (DTFT) is used to compute the spectral density

$$R_{\boldsymbol{\psi}}(t_\ell) = \mathbb{E}\left\{\boldsymbol{\psi}_{\mathrm{f}}(t_k)\boldsymbol{\psi}_{\mathrm{f}}^{\top}(t_k+t_\ell)\right\} \tag{C.10a}$$

$$\Phi^{\mathrm{d}}_{\boldsymbol{\psi}}(\omega) = T\sum_{\ell=-\infty}^{\infty} R_{\boldsymbol{\psi}}(t_\ell)\mathrm{e}^{-j\omega t_\ell}. \tag{C.10b}$$

Following (C.3) and (C.10), $\Phi^{\mathrm{d}}_{\boldsymbol{\psi}}(\omega)$ is linked to $\Phi_{\boldsymbol{\psi}}(\omega)$ in the form

$$\begin{aligned}\Phi^{\mathrm{d}}_{\boldsymbol{\psi}}(\omega) &= T\sum_{\ell=-\infty}^{\infty}\frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi_{\boldsymbol{\psi}}(s)\mathrm{e}^{-jst_\ell}\mathrm{d}s\ \mathrm{e}^{-j\omega t_\ell} \\ &= \int_{-\infty}^{\infty}\Phi_{\boldsymbol{\psi}}(s)\left(\frac{T}{2\pi}\sum_{\ell=-\infty}^{\infty}\mathrm{e}^{j(s-\omega)t_\ell}\right)\mathrm{d}s \\ &= \int_{-\infty}^{\infty}\Phi_{\boldsymbol{\psi}}(s)\sum_{n=-\infty}^{\infty}\delta\left(s-\omega-\frac{2\pi n}{T}\right)\mathrm{d}s \\ &= \sum_{n=-\infty}^{\infty}\Phi_{\boldsymbol{\psi}}\left(\omega+\frac{2\pi n}{T}\right).\end{aligned} \tag{C.11}$$

Then, application of the inverse DTFT to compute $R_{\boldsymbol{\psi}}(t_\ell)$ at $t_\ell = 0$ results in

$$\begin{aligned}R_{\boldsymbol{\psi}}(0) &= \frac{1}{2\pi}\int_{-\pi/T}^{\pi/T}\Phi^{\mathrm{d}}_{\boldsymbol{\psi}}(\omega)\mathrm{d}\omega \\ &= \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\int_{-\pi/T}^{\pi/T}\Phi_{\boldsymbol{\psi}}\left(\omega+\frac{2\pi n}{T}\right)\mathrm{d}\omega.\end{aligned} \tag{C.12}$$

The right-hand side of the above equation is the sum of the matrices defined in (C.7), which is clearly positive definite if $L \geqslant (n_a+n_b+1)/2$. Since the measurement noise is uncorrelated with the input (imposed in Assumption 3), when $\{\boldsymbol{\theta}, \tau\} = \{\boldsymbol{\theta}_{\mathrm{o}}, \tau_{\mathrm{o}}\}$

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N+1}\sum_{k=0}^{N}\boldsymbol{\psi}_{\mathrm{f}}(t_k)\boldsymbol{\phi}_{\mathrm{f}}^{\top}(t_k) &= \lim_{N\to\infty}\frac{1}{N+1}\sum_{k=0}^{N}\boldsymbol{\psi}_{\mathrm{f}}(t_k)\boldsymbol{\psi}_{\mathrm{f}}^{\top}(t_k) \\ &= R_{\boldsymbol{\psi}}(0)\end{aligned} \tag{C.13}$$

is also non-singular. □

References

Ahmed, S. (2020). Step response-based identification of fractional order time delay models. *Circuits, Systems, and Signal Processing*, *39*, 3858–3874.

Baysse, A., Carrillo, F. J., & Habbadi, A. (2011). Time domain identification of continuous-time systems with time delay using output error method from sampled data. In *18th World IFAC congress*.

Björklund, S. (2003). *A survey and comparison of time-delay estimation methods in linear systems* (Ph.D. thesis), Linköping University, Sweden.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis forecasting and control*. San Francisco: Holden-Day.

Chen, F., Garnier, H., & Gilson, M. (2015). Robust identification of continuous-time models with arbitrary time-delay from irregularly sampled data. *Journal of Process Control*, *25*, 19–27.

Chen, F., Garnier, H., Padilla, A., & Gilson, M. (2020). Recursive IV identification of continuous-time models with time delay from sampled data. *IEEE Transactions on Control Systems Technology*, *28*(3), 1074–1082.

Chen, F., Zhuan, X., Garnier, H., & Gilson, M. (2018). Issues in separable identification of continuous-time models with time-delay. *Automatica*, *94*, 258–273.

Ferretti, G., Maffezzoni, C., & Scattolini, R. (1996). On the identifiability of the time delay with least-squares methods. *Automatica*, *32*(3), 449–453.

Garnier, H., & Young, P. C. (2014). The advantages of directly identifying continuous-time transfer function models in practical applications. *International Journal of Control*, *87*(7), 1319–1338.

Gillberg, J., & Ljung, L. (2010). Frequency domain identification of continuous-time output error models, Part II: Non-uniformly sampled data and B-spline output approximation. *Automatica*, *46*, 11–18.

Ha, H., Welsh, J. S., & Alamir, M. (2018). Useful redundancy in parameter and time delay estimation for continuous-time models. *Automatica*, *95*, 455–462.

Kurz, H., & Goedecke, W. (1981). Digital parameter-adaptive control of processes with unknown dead time. *Automatica*, *17*(1), 245–252.

Ljung, L. (1999). *System identification—Theory for the user*. Upper Saddle River: Prentice-Hall.

Ljung, L. (2002). Identification for control: Simple process models. In *41st IEEE conference on decision and control*.

Maruta, I., & Sugie, T. (2013). Projection-based identification algorithm for grey-box continuous-time models. *Systems & Control Letters*, *62*(11), 1090–1097.

Pan, S., González, R. A., Welsh, J. S., & Rojas, C. R. (2020). Consistency analysis of the simplified refined instrumental variable method for continuous-time systems. *Automatica*, *113*, Article 108767.

Pierce, D. A. (1972). Least squares estimation in dynamic disturbance time-series models. *Biometrika*, *5*, 73–78.

Pupeikis, R. (2020). Tracking coefficients of a nonstationary system, followed by static nonlinearity jointly with the time delay. *International Journal of Adaptive Control Signal Processing*, 1–18.

Sánchez, C. J., & Yuz, J. I. (2019). On the relationship between spline interpolation, sampling zeros and numerical integration in sampled-data models. *Systems & Control Letters*, *128*, 1–8.

Söderström, T., & Stoica, P. (1983). *Instrumental variable methods for system identification*. New York: Springer-Verlag.

Solo, V. (1980). Some aspects of recursive parameter estimation. *International Journal of Control*, *32*, 395–410.

Yang, Z., Iemura, H., Kanae, S., & Wada, K. (2007). Identification of continuous-time systems with multiple unknown time delays by global nonlinear least-squares and instrumental variable methods. *Automatica*, *43*(7), 1257–1264.

Young, P. C. (2011). *Recursive estimation and time-series analysis: An introduction for the student and practitioner* (2nd ed.). Berlin: Springer-Verlag.

Young, P. C. (2014). Comment on 'projection-based identification algorithm for grey-box continuous-time models'. *Systems & Control Letters*, *69*, 62–64.

Young, P. C. (2015). Refined instrumental variable estimation: maximum likelihood optimization of a unified Box–Jenkins model. *Automatica*, *51*(1), 35–46.

Young, P. C., & Chen, F. (2021). *A simple robust method of fractional time delay estimation for linear dynamic systems: practical examples*: *Technical note*, Lancaster Environment Centre and Lancaster Data Science Institute, (available from 2nd author).

Young, P. C., & Garnier, H. (2006). Identification and estimation of continuous-time data-based mechanistic (DBM) models for environmental systems. *Environmental Modelling and Software*, *21*(8), 1055–1072.

Young, P. C., Garnier, H., & Gilson, M. (2008). Refined instrumental variable identification of continuous-time hybrid Box–Jenkins models. In H. Garnier, & L. Wang (Eds.), *Identification of continuous-time models from sampled data* (pp. 91–132). London: Springer-Verlag.

Zheng, W. X., & Feng, C. B. (1990). Identification of stochastic time lag systems in the presence of colored noise. *Automatica*, *26*(4), 769–779.

Fengwei Chen was born in Chongqing, China. He received the B.Eng. degree in automation and the M.Eng. degree in control theory and control engineering from Wuhan University, Wuhan, China, in 2009 and 2011, respectively, and the Ph.D. degree in automatic control from the Université de Lorraine, Nancy, France, in 2014.

From 2015 to 2016, he was a Lecturer with the Dalian University of Technology, Dalian, China. From 2017 to 2020, he was an Associate Researcher with Wuhan University, Wuhan, China. Since 2021, he has been with Chongqing University, Chongqing, China, where he is currently an Associate Professor. His research interests include system identification and parameter estimation, with applications to wireless power transfer.

Peter C. Young received the B.Tech. and M.Sc. degrees in aeronautical engineering from Loughborough University, U.K., in 1962 and 1965, respectively, and the M.A. and Ph.D. degrees in automatic control and systems engineering from the University of Cambridge, U.K., in 1970.

From 1968 to 1970, he carried out research on self-adaptive control for the US Navy in California. In 1970, he was appointed as a Lecturer in the Engineering Department and fellow of Clare Hall, University of Cambridge. From 1975 to 1981, he was a Research Professor at the Australian National University in Canberra, Australia, where he helped to establish the Centre for Resource and Environmental Studies. From 1981 to 1987, he was the Head of the Environmental Science Department at Lancaster University, where he was instrumental in establishing the Institute of Environmental and Biological Sciences, and is currently an Emeritus Professor of Environmental Systems. In 1992, he helped to set up the Lancaster Centre for Forecasting with Prof. Robert Fildes and then directed the Centre for Research on Environmental Systems and Statistics at Lancaster. He has published over 300 papers and chapters in books; and authored or coauthored several books, including Recursive Estimation and Time Series Analysis (2011) and True Digital Control (2013). He has also been instrumental, over many years, in the development of CAPTAIN, a noncommercial MATLAB toolbox for the recursive estimation, forecasting and control of discrete and continuous-time dynamic systems. He continues to pursue research on these topics, with applications in diverse areas of research and development.