Journal of Process Control 23 (2013) 1597–1623

Contents lists available at ScienceDirect

Journal of Process Control

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

Review

A tutorial review on process identification from step or relay feedback test

Tao Liu[a,*], Qing-Guo Wang[b], Hsiao-Ping Huang[c]

[a] Institute of Advanced Control Technology, Dalian University of Technology, Dalian, PR China
[b] Department of Electrical & Computer Engineering, National University of Singapore, Singapore
[c] Department of Chemical Engineering, National Taiwan University, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 19 November 2012
Received in revised form 30 July 2013
Accepted 5 August 2013
Available online 13 September 2013

Keywords:
Model identification
Step response test
Relay feedback test
Time delay model
Frequency response estimation
Least-squares fitting

ABSTRACT

Step and relay feedback tests have been widely used for model identification in the process industry. The corresponding identification methods developed in the past three decades are surveyed in this paper. Firstly, the process models with time delay mainly adopted for identification in the literature are presented with a classification on different response types. By categorizing the major technical routes developed in the existing references for parameter estimation relating to different applications, the identification methods are subsequently clustered into groups for overview, along with two specific categories for robust identification against load disturbance and the identification of multivariable or nonlinear processes. The rationales of each category are briefly explained, while a typical or state-of-the-art identification algorithm of each category is elucidated along with application to benchmark examples from the literature to illustrate the achievable accuracy and robustness, for the purpose of facilitating the readers to have a general knowledge of the research development. Finally, an outlook on the open issues regarding step or relay identification is provided to call attention to future exploration.

Contents

* Corresponding author. Tel.: +86 411 84706465; fax: +86 411 84706706.
E-mail addresses: liurouter@ieee.org (T. Liu), elewqg@nus.edu.sg (Q.-G. Wang), huanghpc@ntu.edu.tw (H.-P. Huang).

0959-1524/$ – see front matter
http://dx.doi.org/10.1016/j.jprocont.2013.08.003

Nomenclature

$\mathbb{C}$ field of complex numbers
$\Re_+$ field of nonnegative real numbers
$y(t)$ output response in the time domain
$\Delta y(t)$ output increment to the input change
$\dot{y}(t)$ the first derivative of $y(t)$ in the time domain
$\ddot{y}(t)$ the second derivative of $y(t)$ in the time domain
$\hat{y}(t)$ measured output response in the time domain
$\widehat{y}(t)$ model response in the time domain
$\zeta(t)$ measurement noise in the time domain
$u(t)$ process input in the time domain
$Y(s)$ Laplace transform of output response in the frequency domain
$\Delta Y(s)$ Laplace transform of $\Delta y(t)$
$\hat{Y}(s)$ Laplace transform of $\hat{y}(t)$
$\xi(s)$ Laplace transform of $\zeta(t)$
$U(s)$ Laplace transform of process input in the frequency domain
$L[g(t)]$ Laplace transform of the time domain function, $g(t)$
$G(s)$ process transfer function
$\widehat{G}(s)$ model transfer function
$\mathrm{Re}(G)$ real part of $G \in \mathbb{C}$
$\mathrm{Im}(G)$ imaginary part of $G \in \mathbb{C}$
$\angle G$ phase angle of $G \in \mathbb{C}$
$s=\alpha+j\omega$ Laplace operator, α is the real part (damping factor) and ω is the imaginary part (frequency)
λ time scale factor
T_s sampling period
t_{set} the settling time of a step response
t_N the time corresponding to the Nth sampled data
k_{p} process static gain
τ_{p} process time constant
θ process time delay
M_s number of step tests
ω_{c} cutoff angular frequency
ω_{rc} referential cutoff angular frequency
ω_{p} natural angular frequency of the plant
ω_{u} oscillation angular frequency of the limit cycle
σ_ζ^2 measurement noise variance
u_+ (or u_-) positive (or negative) relay magnitude
ε_+ (or ε_-) positive (or negative) relay switch hysteresis
A_+ (or A_-) positive (or negative) output amplitude in the limit cycle
P_+ (or P_-) positive (or negative) half period corresponding to u_+ (or u_-) in the limit cycle
P_{u} time period of the limit cycle, i.e. $P_{\mathrm{u}}=P_+ + P_- = 2\pi/\omega_{\mathrm{u}}$
A^T transpose of a vector or matrix A
A^* complex conjugate transpose of a matrix A
A^{-1} inverse of a matrix A
$A^\dagger$ pseudo inverse of a matrix A
$\int_{[0,t]}^{(m)} f(t)$ m-fold integrals for a time function, $f(t)$
$F^{(n)}(s)$ the nth order derivative of a complex function, $F(s)$, with respect to s
2DOF two-degree-of-freedom
CSTR continuous stirred tank reactor
DF describing function
Err time domain fitting error
ERR frequency domain fitting error
FFT fast Fourier transform
FOPDT first-order-plus-dead-time
IMC internal model control
IAE integral-of-absolute-error
ISE integral-of-squared-error
IV instrumental variables
LHP left-half-plane
LS least-squares
MIMO multiple-input-multiple-output
MPC model predictive control

NSR	noise-to-signal ratio
P (or PI)	proportional (or proportional-integral)
PID	proportional-integral-derivative
PRBS	pseudo-random binary signal
RHP	right-half-plane
SISO	single-input-single-output
SNR	signal-to-noise ratio
SOPDT	second-order-plus-dead-time

1. Introduction

Control-oriented model identification methods have been increasingly explored in the recent years, owing to the fact that model-based control strategies have been widely recognized for obtaining superior system performance in the setpoint tracking and load disturbance rejection for various industrial and chemical processes. Among a variety of excitation signals adopted for model identification, the step test is most widely practiced given its simple and economic implementation. Generally, a step test is performed in terms of an open-loop structure and initiated when the process to be identified is at zero initial state or moved into the desired operating region (see Fig. 1), especially around the operating point, namely, the setpoint. To prevent the process output from drifting too far away from the setpoint, as required from the operation of many industrial processes such as the continuous stirred tank reactors (CSTRs) and heating boilers, a closed-loop identification test is usually adopted in engineering practice to keep the output deviation in an admissible working range. There are two types of closed-loop identification test that are widely used in engineering applications, closed-loop step test (see Fig. 2) and relay feedback test (see Fig. 3). For the use of a closed-loop step test, the closed-loop controller needs to be specified beforehand for maintaining the closed-loop stability, which may bring difficulty to the closed-loop configuration since the process is to be identified. Hence, a closed-loop step test is usually preferred to online identification in order to improve the controller tuning in practical applications. In contrast, no prior knowledge of the closed-loop structure is required to perform a relay feedback test. Closed-loop identification methods based on relay feedback tests have therefore been developed on an ad hoc basis in the past three decades. The pioneering work can be found in the literature [1–4]. Recent monographs concerned with relay identification can be seen in [5–8].

As surveyed in the early literature [9], earlier references on step identification had been mainly devoted to identifying linear models free of time delay. In fact, time delay is usually associated with industrial process operations regarding the execution of the setpoint command, mass transportation, and output/signal transmission etc., which is usually indicated by the output response delay observed from the sampled data. The phenomenon may deteriorate severely the control performance or even cause the control system unstable [10,11]. To facilitate the control system design and online tuning in practice, the research efforts in the past three decades have therefore been devoted mainly to the identification of linear low-order models plus time delay such as the first-order-plus-dead-time (FOPDT) and second-order-plus-dead-time (SOPDT). Note that a low-order model of FOPDT or SOPDT can also be effectively used to describe the fundamental dynamics of a higher-order linear process for the purpose of controller tuning [5–8,11,12]. Owing to the fact that the process response characteristics in the low frequency range corresponding to a phase change from zero to $-\pi$ are primarily concerned for controller tuning in engineering practice [4–8,11–17], identification accuracy of these low-order models has been mostly evaluated in terms of the fitting accuracy in the low frequency range including the zero frequency that corresponds to the static gain of the process transfer function. Nevertheless, with the wider application of model predictive control (MPC) methods using modern computer technology [11], the development of higher-order model identification methods using step or relay tests to achieve a better fit over a wider frequency range [7,8] is required in the recent years.

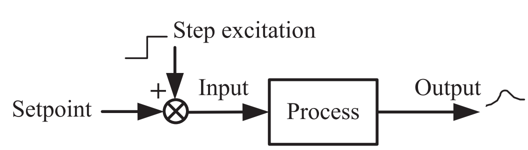

Fig. 1. Schematic of an open-loop step test.

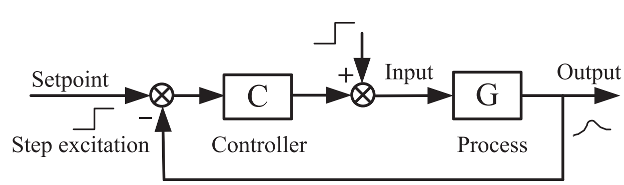

Fig. 2. Schematic of a closed-loop step test.

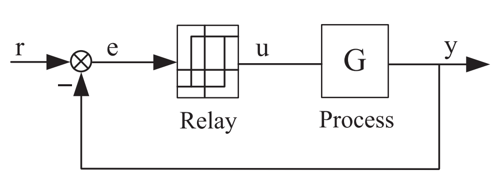

Fig. 3. Schematic of a relay feedback test.

For the use of relay feedback test, the historical development of the describing function (DF) approach was reviewed in the literature [13], together with a summary of the guidelines for tuning the conventional proportional-integral (PI) or proportional-integral-derivative (PID) controller based on the limit cycle information in the steady oscillations. Besides, the literature [14] gave a review on using the transient response in a relay test and parasitic relay loops to obtain more process frequency response information for improving the PID controller tuning, based on previous results explored by the authors.

Although a large amount of research result has been cultivated for model identification from step or relay tests in the past three decades, no overview, as far as we know, has been reported as yet on the overall research development of either step or relay identification methods. Based on a series of research results explored in the past years, some of which have been summarized in the monographs [5,8], we are motivated to give a tutorial review on the research development of different identification approaches based on using step or relay feedback tests. To have a clear understanding in reviewing the existing references, we make a classification on the technical routes adopted for developing step or relay identification methods, such that the corresponding references are clustered into groups for overview on the rationales. Concerning the identification of linearized process models, the main technical routes of the existing step identification methods are categorized into three groups, one is the use of representative points in the transient response for model fitting, another is the time integral approach, and the third is the use of frequency response estimation for model fitting. Also, the main technical routes of the existing relay identification methods are categorized into three groups, one is the describing function (DF) method, another is the curve fitting approach, and the third is the use of frequency response estimation for model fitting. For each category, a typical or state-of-the-art identification algorithm will be presented with an explanation on the rationale for the ease of understanding. Moreover, robust identification methods based on using step or relay feedback tests to cope with the influence from load disturbance, measurement noise, and/or nonzero (or unsteady) initial process conditions that are often encountered in practical applications are separately clustered for clarity. In addition, some papers reported the use of step or relay feedback test to identify specific multiple-input-multiple-output (MIMO) or nonlinear processes, and therefore, are classified into additional categories for review.

In this survey, all the published papers were found via literature search in the control-related journals collected in the worldwide electrical databases available to the authors, including ScienceDirect, IEEE Xplore, ACS, Wiley, and Taylor & Francis. Three searching fields, Abstract, Title and Keywords, were chosen, and the searching phrases were taken as 'step' and 'relay' in combination with 'identification', within the time span from Jan 1980 to Oct 2012. We shall clarify that some papers relating to controller autotuning or performance monitoring also presented step or relay identification methods, some of which are cited here based on our study in the past years, but the rest are not cited due to the difficulty in finding them out from enormous references with a title word of 'autotuning' or 'monitoring'. Other relevant papers might also be overlooked due to mismatch with the above searching conditions. Hopefully, the major references cited herein, together with the 'mismatched' but closely related papers we have studied in the past years, should have included the main contributions developed in the past three decades, from our viewpoint and that of our research collaborators.

The paper objectives are twofold, one is for the readers to have a general knowledge of the research development on step or relay identification methods, and another is for calling attention to the unexplored research topics relating to these two experimental approaches together with some challenging issues that are left open as yet but solicited extensively in industrial applications. For clarity, the paper is organized as follows: Section 2 presents the model structures mainly adopted in the existing literature for step or relay identification. In Section 3 a survey on the existing step identification methods is presented based on a classification on the technical routes adopted for model identification together with a specific category for robust identification against nonzero (or unsteady) initial process conditions and/or load disturbance. Illustrations using benchmark examples from the literature are given to demonstrate the achievable identification accuracy and robustness of the categorized identification methods. Subsequently, another survey on the existing relay identification methods is given with a similar categorization in Section 4 together with illustrative examples. A brief review on step or relay identification of nonlinear processes based on a Hammerstein- or Wiener-type model structure is presented in Section 5. Finally, some concluding remarks and an outlook on prospective research topics are addressed in Section 6.

2. Model types for identification

In the literature relating to step or relay identification published in the past three decades, low-order linear models plus time delay have been mainly adopted for identification owing to the demands from practical applications for the model-based control design and tuning. The primary concern of identifying such a model is the fitting accuracy in the low frequency range as aforementioned that reflects the fundamental dynamic response characteristics of the plant or the closed-loop system for operation. The commonly used model types are FOPDT, SOPDT, and higher-order linear model plus time delay.

Generally, an FOPDT model is written in the form of

$$G_1(s) = \frac{k_p}{\tau_p s \pm 1} e^{-\theta s} \tag{1}$$

where k_p denotes the process gain, θ the process time delay, and τ_p the process time constant. The sign '+' corresponds to a stable process, and in the opposite, '−' indicates an unstable process. Note that if $\tau_p \to \infty$, the above model may be simplified to $G_{1-I}(s) = (k_p/\tau_p)e^{-\theta s}/s$, which represents an integrating process.

For an SOPDT model, there are three forms commonly studied in the literature,

$$G_{2-S}(s) = \frac{k_p}{\tau_p^2 s^2 + 2\xi\tau_p s + 1} e^{-\theta s} \tag{2}$$

$$G_{2-I}(s) = \frac{k_p}{s(\tau_p s + 1)} e^{-\theta s} \tag{3}$$

$$G_{2-U}(s) = \frac{k_p}{(\tau_1 s - 1)(\tau_2 s + 1)} e^{-\theta s} \tag{4}$$

where ξ in the stable process model is customarily named as the process damping ratio, while τ_p is a time constant equal to the reciprocal of ω_p that is customarily named as the natural angular frequency, according to a classification on the process response types, i.e. stable, integrating and unstable, in engineering practice. In (3) and (4), τ_p, τ_1 and τ_2 are positive constants while τ_p (or τ_2) reflects the inertia character (response speed) of an integrating (or unstable) process.

Based on a classification on the range of ξ, three types of a stable process model have been widely used, underdamped ($0 < \xi < 1$), critically damped ($\xi = 1$), and overdamped ($\xi > 1$), which are also written in the following forms for identification, respectively,

$$G_{2-\mathrm{SO}}(s) = \frac{k_p}{a_2 s^2 + a_1 s + 1} e^{-\theta s} \quad (5)$$

$$G_{2-\mathrm{SC}}(s) = \frac{k_p}{(\tau_p s + 1)^2} e^{-\theta s} \quad (6)$$

$$G_{2-\mathrm{SU}}(s) = \frac{k_p}{(\tau_1 s + 1)(\tau_2 s + 1)} e^{-\theta s} \quad (7)$$

where a_1, a_2, τ_p, τ_1 and τ_2 are positive coefficients.

A higher-order model plus time delay is generally written by

$$G_n(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \cdots b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \cdots a_1 s + 1} e^{-\theta s} \quad (8)$$

where b_0 denotes the process static gain, i.e., k_p and $n > m$ indicates strict properness of the process transfer function.

3. Step identification methods

Generally, an open-loop step test is performed when the process is at zero initial state or in a nonzero steady state, such that the obvious dynamic (or transient) response of the process to a step change of the process input is observed for model identification. It is obvious that a larger magnitude of the step change can facilitate a better observation of the transient response. This, however, is subject to the process operation constraints in practice. Most existing step identification methods have been developed based on using zero initial conditions or nonzero steady state to perform a step test.

In contrast, when a closed-loop step test is used for model identification, the step change is usually added to the setpoint rather than the process input because any external signal added to the process input acts like a load disturbance that may be rejected by the closed-loop feedback mechanism. Also, a closed-loop step test is generally performed after the closed-loop system has already moved into a steady state. To facilitate model identification from a closed-loop step test, the closed-loop controller should be prescribed in a simple form like the proportional (P), PI, or PID type, such that an analytical or quantitative relationship between the process response and the closed-loop system response can be explicitly constructed [8,15].

Based on a classification on the main technical routes adopted in the existing references as cited, we give a tutorial review following the chronological occurrence of these identification methods in the following subsections.

3.1. Using representative points in the transient response for model fitting

Since the two-point fitting idea based on the transient step response was proposed for a rough estimate of an FOPDT model, as surveyed in the early literature [9], a widely recognized two-point fitting algorithm for obtaining an FOPDT model had been summarized by Åström and Hägglund [12]. In brief, the static gain k_p for a stable process is obtained from the ratio of the steady-state output deviation over the step change, and the time delay θ is read as the intercept of the tangent to the step response that has the steepest slope with respect to the time axis, as shown in Fig. 4. The time constant τ_p is determined as the difference between θ and the time spent for the step response to reach a value of $0.63k_p$. Another recognized two-point fitting method was reported by Marlin [16], that is, choosing the time points at which the step response reaches 28% and 63% of the steady-state output deviation, respectively, for the computation of τ_p and θ. By establishing the fitting conditions between these representative points and the time domain step response expression, alternative parameter estimation formulae were presented in the literature [11]. This approach was extended to obtain a higher-order model with time delay by using the tangent information of the inflection point in the transient response [17], which is effective mainly for industrial stable processes with underdamped step response.

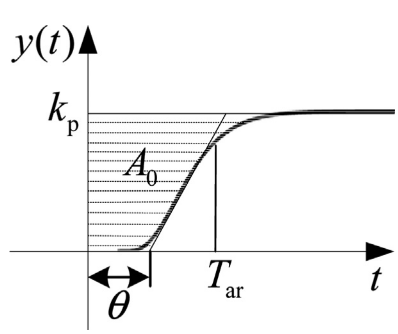

Fig. 4. Graphical determination of the model parameters from an open-loop step test.

Table 1
Illustrative comparison between the categorized step identification methods.

Process	NSR (%)	Two- or three-point fitting	Time integral approach	Frequency response fitting
$G_1=\frac{1}{s+1}e^{-s}$	0	$\frac{1}{0.9s+1}e^{-1.1s}$ (Ref. [12])	$\frac{1.0}{0.997s+1}e^{-1.0s}$ (Ref. [22])	$\frac{0.9999}{0.9999s+1}e^{-1.0001s}$ (Ref. [32])
$G_2=\frac{1.25e^{-0.234s}}{0.25s^2+0.7s+1}$	0	$\frac{1.25e^{-0.2595s}}{0.2433s^2+0.6856s+1}$ (Ref. [20])	$\frac{1.2501e^{-0.2339s}}{0.2499s^2+0.6999s+1}$ (Ref. [26])	$\frac{1.2499e^{-0.2341s}}{0.2501s^2+0.6999s+1}$ (Ref. [32])
	10	–	$\frac{1.2503(\pm0.0231)e^{-0.22(\pm0.034)s}}{0.258(\pm0.028)s^2+0.713(\pm0.041)s+1}$ (Ref. [39])	$\frac{1.2513(\pm0.016)e^{-0.234(\pm0.04)s}}{0.249(\pm0.03)s^2+0.71(\pm0.03)s+1}$ (Ref. [32])
$G_3=\frac{2.15(-2.7s+1)(158.5s^2+6s+1)}{(17.5s+1)^4(20s+1)}e^{-14s}$	0	$\frac{2.15e^{-21.56s}}{2234.67s^2+71.99s+1}$ (Ref. [19])	$\frac{2.15e^{-28.58s}}{1939.41s^2+69.63s+1}$ (Ref. [26])	$\frac{2.1413e^{-27.96s}}{1903.12s^2+70.975s+1}$ (Ref. [32])
	10	–	$\frac{2.149(\pm0.041)e^{-28.63(\pm0.054)s}}{1938.76(\pm0.035)s^2+69.84(\pm0.047)s+1}$ (Ref. [26])	$\frac{2.147(\pm0.032)e^{-27.98(\pm0.061)s}}{1906.45(\pm0.039)s^2+70.46(\pm0.041)s+1}$ (Ref. [32])
Computation effort		Low	Moderate or relatively high	Moderate

Where Ref. [·] in the parentheses of each model denotes the cited reference presenting the identification algorithm; For the Monte Carlo tests in the presence of measurement noise, the identified models are shown by the mean value of each parameter along with the sample standard deviation in the parentheses.

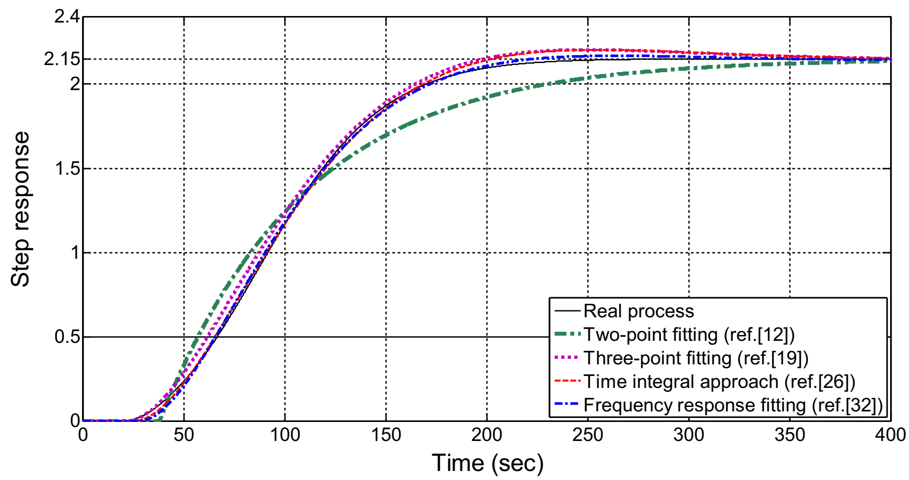

Fig. 5. Illustrative comparison of the categorized step identification methods for fitting the step response of a high-order process.

When an oscillatory output response is observed under a step test, a three-point fitting method was proposed by Huang and Clements [18] to identify an SOPDT model. The objective was to locate the most suitable three intersection points that could minimize the integral-of-absolute-error (IAE) criterion with respect to the transient response difference between the process and the identified model. By deriving the analytical time response expressions and the extreme value points based on an SOPDT model response, improved three-point fitting algorithms for better accuracy were developed in Refs. [19–21].

An obvious merit of the above two- and three-point fitting methods is the simplicity for implementation. For illustration, the two- and three-point fitting algorithms in Refs. [12,19,20] are used to identify the first-, second-, and fifth-order processes (G_1, G_2, and G_3) listed in Table 1. The step response tests for model identification are performed the same as those detailed in Refs. [22,26,32], respectively. A comparison on fitting the step response of G_3 is shown in Fig. 5. To illustrate the achievable control performance for G_3, the standard internal model control (IMC) method [28] is adopted based on the identified FOPDT and SOPDT models listed in Table 1. By adding a step change of the setpoint and assuming an step-type load disturbance with a magnitude of −0.5 that occurs at $t = 500(s)$ from the process input side, the control results are shown in Fig. 6, where the IMC tuning parameter is taken to obtain the similar rising speed of the setpoint tracking

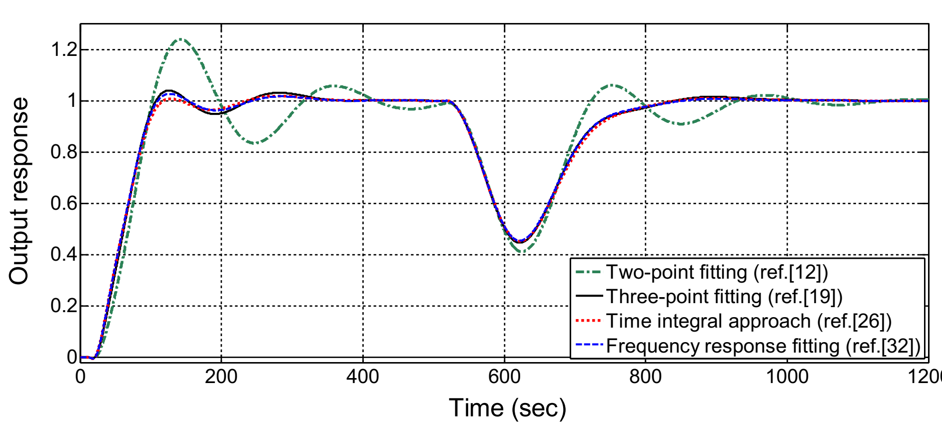

Fig. 6. Illustrative comparison of the categorized step identification methods for the IMC performance on a high-order process.

Table 2
Comparison of ISE for the setpoint tracking and load disturbance by using the standard IMC based on the categorized step identification methods.

Identification methods	ISE	
	Setpoint tracking	Disturbance attenuation
Two-point fitting (Ref. [12])	59.4577	28.1302
Three-point fitting (Ref. [19])	49.8591	26.1661
Time integral approach (Ref. [26])	48.8024	26.1582
Frequency response fitting (Ref. [32])	48.6681	25.5527

for comparison between using these identified models. The corresponding integral-of-squared errors (ISEs) for the setpoint tracking and load disturbance rejection are listed in Table 2.

It should be noted that such a two- or three-point fitting method may be quite sensitive to measurement noise that is usually confronted in engineering applications, possibly causing unacceptable identification errors for the subsequent model-based control design. As can be verified from the applications to the noisy tests indicated by the noise-to-signal ratio (NSR) in Table 1, it is really difficult to locate the representative points required in the referred two- or three-point fitting algorithm to obtain a desirable model, which is therefore omitted.

3.2. Time integral approach

Motivated by improving the identification accuracy and robustness against measurement noise, the time integral approach was developed in the 1990s. A pioneering work using the time integral to compute the graphical area regarding the transient step response for model identification was explored by Åström and Hägglund [12]. Consider the identification of an FOPDT model for an overdamped stable process based on the step response shown in Fig. 4. The static gain is determined by a ratio of the steady-state output deviation over the step change. The average residence time, $T_{\rm ar}$, is computed from the area A_0 shown in Fig. 4 as

$$T_{\rm ar} = \frac{A_0}{k_{\rm p}} = \frac{\int_0^{\infty}[\Delta y(\infty) - \Delta y(t)]dt}{k_{\rm p}} \tag{9}$$

where $\Delta y(t)$ denotes the output response to the step change.

The time constant and time delay parameters can then be determined by

$$\begin{cases} \tau_{\rm p} = \dfrac{eA_1}{k_{\rm p}} = \dfrac{e\displaystyle\int_0^{T_{\rm ar}} \Delta y(t)dt}{k_{\rm p}} \\ \theta = T_{\rm ar} - \tau_{\rm p} \end{cases} \tag{10}$$

where 'e' is a constant, i.e. a base of the exponential function.

Enhanced fitting accuracy for identifying an FOPDT model was proposed by Bi et al. [22], based on using a least-squares (LS) fitting algorithm. The rationale is briefed as below.

Assume that the initial conditions of a process to be identified are set to zero. The time domain output response to a step change of the input, $u(t) = h$, can be written in terms of an FOPDT model in (1) as

$$\tau_{\rm p}\dot{y}(t) + y(t) = k_{\rm p}u(t-\theta), \quad t \geq \theta \tag{11}$$

where $y(t) = 0$ for $t < \theta$ is assumed.

Integrating both sides of (11) for $t < t_N$, where N is the sampled data length chosen in terms of the transient response, yields

$$\int_0^t y(\tau)d\tau = k_{\rm p}h(t-\theta) - \tau_{\rm p}y(t) \tag{12}$$

which can be reformulated as

$$\psi(t) = \phi^T(t)\gamma \tag{13}$$

where

$$\begin{cases} \psi(t) = \displaystyle\int_0^t y(\tau)d\tau, \\ \phi(t) = [ht, -h, -y(t)]^T, \\ \gamma = [k_{\rm p}, k_{\rm p}\theta, \tau_{\rm p}]^T. \end{cases} \tag{14}$$

Using the sampled transient response data, $\theta < t_1 < t_2 < \cdots < t_N$, the following data matrices can be constructed,

$$\Psi = [\psi(t_1), \psi(t_2), \cdots, \psi(t_N)]^T \tag{15}$$

$$\Phi = [\phi(t_1), \phi(t_2), \cdots, \phi(t_N)]^T \tag{16}$$

Hence, an LS algorithm can be established for parameter estimation,

$$\Psi = \Phi\gamma \tag{17}$$

where the parameter vector can be solved by

$$\gamma = (\Phi^T \Phi)^{-1} \Phi^T \Psi \tag{18}$$

Accordingly, the FOPDT model parameters can be retrieved by

$$\begin{bmatrix} k_{\rm p} \\ \theta \\ \tau_{\rm p} \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \gamma(2)/\gamma(1) \\ \gamma(3) \end{bmatrix} \tag{19}$$

In the presence of measurement noise, $\zeta(t)$, there exists $\hat{y}(t) = y(t) + \zeta(t)$, where $\hat{y}(t)$ denotes the actually measured output value. Substituting it into (11) and then integrating both sides of (11) yields

$$\psi(t) = \phi^{\rm T}(t)\gamma + \nu(t) \tag{20}$$

where $\psi(t) = \int_0^t \hat{y}(\tau)d\tau$, $\phi(t) = \left[ht, -h, -\hat{y}(t)\right]^T$, γ is the same as in (14), and $\nu(t) = -\tau_{\rm p}\zeta(t) - \int_0^t \zeta(\tau)d\tau$.

It is seen from (20) that $\phi^T(t)$ is now correlated with $\nu(t)$, i.e., the influence from measurement noise. The parameter estimation from (18) cannot guarantee consistency according to the parameter estimation theory [23,24]. To overcome the problem, the instrumental variable (IV) method [23] can be used, that is, an instrumental matrix, Z, may be selected such that

(1) the inverse of $\lim_{N\to\infty} (Z^T \Phi)/N$ exists;
(2) $\lim_{N\to\infty} (Z^T \upsilon)/N = 0$, where $\upsilon = [\nu(t_1), \nu(t_2), \cdots, \nu(t_N)]^T$.

A consistent parameter estimation is therefore obtained by

$$\gamma = (Z^T \Phi)^{-1} Z^T \Psi \tag{21}$$

A feasible choice of the instrumental matrix was suggested by Bi et al. [22],

$$Z = \begin{bmatrix} t_1 & -1 & \dfrac{1}{t_1} \\ t_2 & -1 & \dfrac{1}{t_2} \\ \vdots & \vdots & \vdots \\ t_N & -1 & \dfrac{1}{t_N} \end{bmatrix} \tag{22}$$

Note that the above identification algorithm is based on using the transient response data with $\theta < t_1 < t_2 < \cdots < t_N$, where θ is to be identified. For practical application, it was suggested [22] to perform a step test in this way: after the process has moved into a zero or nonzero steady state, the process output should be monitored for a short 'listening period', through which the measurement noise band, B_n, may be estimated. Then, the step test is performed until the output response settles down. The output response satisfying

$$\left|\hat{y}(t)\right| > 2B_n \tag{23}$$

or alternatively,

$$\left|\hat{y}(t)\right| > 0.1\hat{y}(t_N) \tag{24}$$

can be collected for application of the above identification algorithm.

Following the above time integral method for identifying an FOPDT model, extended identification algorithms for identifying an SOPDT or higher-order plus time delay model were developed in the subsequent Refs. [25,26]. It was demonstrated by a few benchmark examples that these time integral identification algorithms could significantly improve the identification accuracy and robustness against measurement noise in comparison with previous step identification methods. The computation effort, however, is relatively high. To obtain a nth order process model ($n \geq 2$), n-fold integrals with respect to all the sampled transient response data are adopted in such an identification algorithm for parameter estimation.

For illustration, the identification algorithms in Refs. [22,26] are used to identify the first-, second-, and fifth-order processes (G_1, G_2, and G_3) listed in Table 1 for comparison with the other categorized methods, along with comparisons on fitting the step response of G_3 and the IMC-based control results, which are shown in Figs. 5 and 6, respectively. Note that for the presence of measurement noise with NSR = 10%, 200 Monte Carlo tests are performed to obtain the identification results for G_2 and G_3 as listed in Table 1, each of which is shown by the mean value of model parameter along with the sample standard deviation in the parentheses.

It should be noted that the time integral approach based on an open-loop step test can also be used for the identification of integrating processes, when zero initial process conditions or a reasonable zero initialization of the process states can be obtained for the implementation of a step test, as studied by Liu et al. [27] for model identification and online autotuning of the temperature control system for an industrial injection molding machine.

3.3. Frequency response estimation for model fitting

For the identification of transfer function models, frequency domain identification methods have been subsequently developed based on using step response tests. Generally, a two-step identification procedure is adopted, the first step is to estimate the process frequency

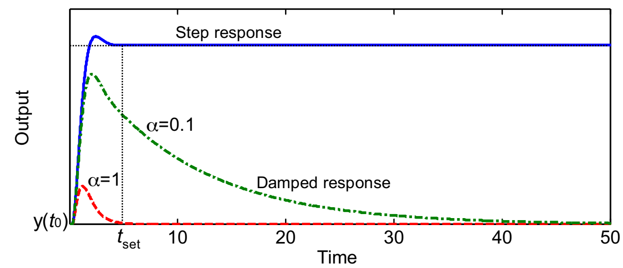

Fig. 7. Illustration of frequency response estimation by choosing the damping factor (α) based on a step test.

response at a few specified frequencies that are usually in the low frequency range with a phase change from zero to $-\pi$ which is primarily concerned for control design and controller tuning [28–31], and the second step is parameter estimation by fitting the estimated frequency response points in terms of a specified model structure such as FOPDT or SOPDT. For illustration, the rationale of the state-of-the-art frequency response estimation method [32] is briefly presented as below.

When a step change is added to the process input for a step test as shown in Fig. 7, it is commonly known that the Fourier transform of the output response does not exist due to $\Delta y(t) \neq 0$ for $t \to \infty$, where $\Delta y(t) = y(t) - y(t_0)$ and $y(t_0)$ denotes the initial output value of a steady state. Nevertheless, by substituting $s = \alpha + j\omega$ into the Laplace transform for the step response,

$$\Delta Y(s) = \int_0^{\infty} \Delta y(t) e^{-st}\, dt \tag{25}$$

one can formulate

$$\Delta Y(\alpha + j\omega) = \int_0^{\infty} [\Delta y(t) e^{-\alpha t}] e^{-j\omega t}\, dt \tag{26}$$

Note that, if $\alpha > 0$, there exists $y(t)e^{-\alpha t} = 0$ for $t > t_N$, where t_N may be numerically determined using the condition that $\Delta y(t_N)e^{-\alpha t_N} \to 0$, since $\Delta y(t)$ reaches a steady value after the transient response.

Therefore, by taking α as a damping factor to the step response for the Laplace transform, one can compute $\Delta Y(\alpha + j\omega)$ from the N points of step response data as

$$\Delta Y(\alpha + j\omega) = \int_0^{t_N} [\Delta y(t) e^{-\alpha t}] e^{-j\omega t}\, dt \tag{27}$$

For a step test with an initial steady state of the process, i.e. $y(t) = c$ for $t \le t_0$, as shown in Fig. 7, by using a time shift of t_0 (i.e. letting $t_0 = 0$), the step change can be written as

$$\Delta u(t) = \begin{cases} 0, & t \le 0; \\ h, & t > 0. \end{cases} \tag{28}$$

where h is the magnitude of the step change. Correspondingly, its Laplace transform for $s = \alpha + j\omega$ with $\alpha > 0$ can be expressed by

$$\Delta U(\alpha + j\omega) = \int_0^{\infty} h e^{-(\alpha + j\omega)t}\, dt = \frac{h}{\alpha + j\omega} \tag{29}$$

Hence, a frequency response estimation of the process can be derived using (27) and (29) as

$$G(\alpha + j\omega) = \frac{\alpha + j\omega}{h} \Delta Y(\alpha + j\omega), \quad \alpha > 0 \tag{30}$$

Note that it is infeasible to evaluate $G(j\omega)$ directly in terms of the fast Fourier transform (FFT) and its inverse transform as introduced in the literature (e.g. [5]), i.e. $G(j\omega_k) = \text{FFT}\{\text{FFT}^{-1}\{G(\alpha + j\omega_k)\}\, e^{\alpha k T_s}\}$, where T_s is the sampling period for computing the numerical integral in (27). The reason lies with the fact that the step response is not periodic signal that can be topologized for FFT computation.

In the presence of measurement noise, $\zeta(t) \sim N(0, \sigma_\zeta^2)$, it was proved [32] that unbiased frequency response estimation can be obtained by

$$\Delta Y(\alpha + j\omega) = \lim_{M_s \to \infty} \frac{1}{M_s} \sum_{i=1}^{M_s} \Delta \hat{Y}_i(\alpha + j\omega) \tag{31}$$

and the estimation error variance is bounded by

$$\lim_{M_s \to \infty} \sigma_e^2 \le \frac{T_s^2 \sigma_\zeta^2}{1 - e^{-2T_s \alpha}} \tag{32}$$

where $\Delta Y(\alpha + j\omega) = L[\Delta y(t)]$, $\Delta \hat{Y}_i(\alpha + j\omega) = L[\Delta \hat{y}_i(t)]$, $\Delta y(t)$ is the true value of the step response, $\Delta \hat{y}_i(t) = \Delta y(t) + \zeta_i(t)$ the measured output response in each test, and M_s the number of step tests.

It is obvious that $G(\alpha+j\omega)\to 0$ as $\alpha\to\infty$. On the contrary, $\alpha\to 0$ will cause t_N to be much larger when computing (30). A proper choice of α is therefore required for computation. Considering that all the transient step response data should be used to ensure good estimation of the process frequency response, the following condition was suggested [32] to choose α,

$$\left|\Delta y(t_{\rm set})\right| T_s e^{-\alpha t_{\rm set}} > \delta \tag{33}$$

where $\Delta y(t_{\rm set})=y(t_{\rm set})-y(0)$ denotes the transient output response to the step change in terms of the settling time ($t_{\rm set}$), in which $y(0)$ indicates the initial output value before the step test, T_s is the sampling period for the computation of the numerical integral in (27), and δ is a computational threshold which can be practically taken smaller than $\left|\Delta y(t_{\rm set})\right| T_s\times 10^{-6}$.

It follows from (33) that

$$\alpha < \frac{1}{t_{\rm set}} \ln \frac{\left|\Delta y(t_{\rm set})\right| T_s}{\delta} \tag{34}$$

Once α is chosen in terms of the above guideline, the time length, t_N, can be determined from a numerical constraint for computing (27),

$$\left|\Delta y(t_N)\right| e^{-\alpha t_N} < \delta \tag{35}$$

which can be solved as

$$t_N > \frac{1}{\alpha} \ln \frac{\left|\Delta y(t_N)\right|}{\delta} \tag{36}$$

Note that for a very slow process in practice, it may happen that any value of α chosen to comply with (34) may also satisfy $\alpha<\delta$, which will deteriorate the efficiency for computing (30). In such a case, it was suggested [32] that a time-scaled output response be used to perform the Laplace transform, i.e. $L[\Delta y(t/\lambda)]=\lambda\Delta Y[\lambda(\alpha+j\omega)]$, where λ is a time scale factor. Therefore, the scaled frequency response, $G[\lambda(\alpha+j\omega)]$, can be effectively computed for model identification.

Based on the estimated frequency response in the first step, the second step of model identification can be proceeded with in terms of a specified model structure. For example, assume that a time delay model with repeated poles is to be identified,

$$\widehat{G}(s) = \frac{k_{\rm p}}{(\tau_{\rm p}s+1)^m} e^{-\theta s} \tag{37}$$

where m denotes the number of repeated poles which also indicates the process order. It is obvious that $m=1$ corresponds to an FOPDT model.

To avoid confusion, here denote the nth order derivative for a complex function of $F(s)$ with respect to s by

$$F^{(n)}(s) = \frac{d^{(n)}}{ds^n} F(s), \quad n\geq 1 \tag{38}$$

It follows from (27) and (30) that

$$G^{(1)}(s) = \frac{1}{h}\int_0^\infty (1-st)\Delta y(t) e^{-st}\,dt \tag{39}$$

$$G^{(2)}(s) = \frac{1}{h}\int_0^\infty t(st-2)\Delta y(t) e^{-st}\,dt \tag{40}$$

Hence, by letting $s=\alpha$ and choosing α as well as that for computing (27), the numerical integrals in (39) and (40) can be computed. The corresponding time lengths of t_N may be determined, respectively, using the numerical constraints,

$$\left|(1-\alpha t_N)\Delta y(t_N)\right| T_s e^{-\alpha t_N} < \delta \tag{41}$$

$$\left|t_N(\alpha t_N-2)\Delta y(t_N)\right| T_s e^{-\alpha t_N} < \delta \tag{42}$$

By viewing $s\in\Re_+$ and taking the natural logarithm for both sides of (37), one obtains

$$\ln[\widehat{G}(s)] = \ln(k_{\rm p}) - m\ln(\tau_{\rm p}s+1) - \theta s \tag{43}$$

Then, taking the first and second derivatives for both sides of (43) with respect to s yields

$$\frac{\widehat{G}^{(1)}(s)}{\widehat{G}(s)} = -\frac{m\tau_{\rm p}}{\tau_{\rm p}s+1} - \theta \tag{44}$$

$$\frac{\widehat{G}^{(2)}(s)\widehat{G}(s) - [\widehat{G}^{(1)}(s)]^2}{\widehat{G}^2(s)} = \frac{m\tau_{\rm p}^2}{(\tau_{\rm p}s+1)^2} \tag{45}$$

For simplicity, denote the left-hand side of (44) as $Q_1(s)$ and the left-hand side of (45) as $Q_2(s)$.

By substituting $s=\alpha$, $\widehat{G}(\alpha)=G(\alpha)$, $\widehat{G}^{(1)}(\alpha)=G^{(1)}(\alpha)$, and $\widehat{G}^{(2)}(\alpha)=G^{(2)}(\alpha)$ into (45), it can be derived that

$$\tau_{\rm p}=\begin{cases}\dfrac{-\alpha Q_2+\sqrt{mQ_2}}{\alpha^2 Q_2-m}, & \text{if} \quad \alpha^2 Q_2-m>0;\\ \dfrac{\alpha Q_2+\sqrt{mQ_2}}{m-\alpha^2 Q_2}, & \text{if} \quad \alpha^2 Q_2-m<0.\end{cases} \tag{46}$$

Consequently, the remaining parameters can be derived from (44) and (37) as

$$\theta=-Q_1(\alpha)-\frac{m\tau_{\rm p}}{\tau_{\rm p}\alpha+1} \tag{47}$$

$$k_{\rm p}=(\tau_{\rm p}\alpha+1)^m G(\alpha)e^{\alpha\theta} \tag{48}$$

Note that the above identification algorithm can ensure identification accuracy if the model structure matches the process to be identified. In case there exists model mismatch when identifying a process with unknown model structure, an accurate fitting may not be guaranteed since the above algorithm establish frequency response fitting only around the zero frequency (i.e. $\omega=0$). To improve the model fitting accuracy over a user-specified frequency range, e.g. the referred low frequency range for control design, the following identification algorithm can be adopted [32] for identifying a model in the general form of (8):

Let $s=\alpha+j\omega_k$ $(k=1, 2, \ldots, M)$, where M is the number of representative frequency response points in the specified frequency range, the objective function for model identification is

$$J_{opt}=\sum_{k=0}^{M}\rho_k\left|G(\alpha+j\omega_{\rm k})-\widehat{G}(\alpha+j\omega_{\rm k})\right|^2<ERR^2 \tag{49}$$

where $G(\alpha+j\omega_{\rm k})$ and $\widehat{G}(\alpha+j\omega_{\rm k})$ denotes the frequency responses of the process and the model, respectively, ERR is a user-specified threshold for assessing the fitting accuracy, and ρ_k $(k=0, 1, 2, \ldots, M)$ are weighting coefficients for emphasizing frequency response fitting over the specified frequency range.

To guarantee identification robustness against measurement noise, one can choose

$$\omega_{\rm M}=(1.0\text{–}2.0)\omega_{\rm rc} \tag{50}$$

$$\rho_k=\eta^k\Big/\sum_{k=0}^{M}\eta^k,\ \eta\in[0.9, 0.99] \tag{51}$$

in consideration of that the step response has inherently a low signal-to-noise ratio (SNR) for the high-frequency part over measurement noise. Note that, if there exists low- or middle-frequency noise, a denoising low-pass filter should be devised based on the knowledge of the process response characteristics and the measurement sensor to exclude the influence from such noise.

Owing to $\sum_{k=0}^{M}\rho_k ERR^2=ERR^2$, the objective function in (49) may accommodate the widely used frequency response error specification [33],

$$ERR=\max_{\omega\in[0,\omega_c]}\left\{\left|\frac{G(j\omega)-\widehat{G}(j\omega)}{G(j\omega)}\right|\right\} \tag{52}$$

where ω_c is the cutoff angular frequency corresponding to $\angle G(j\omega_c)=-\pi$. In practical terms, it is difficult to know ω_c exactly. The above choice of $\omega_{\rm M}$ should cover ω_c for computation.

Substituting (30) and the general model form of (8) into the left-hand side of (49) by letting it equal zero, and then organizing the resulting expression into an LS form of

$$\psi(s)=\phi(s)^T\gamma \tag{53}$$

where

$$\begin{cases}\psi(s)=\Delta Y(s),\\ \phi(s)=[-s^n\Delta Y(s), -s^{n-1}\Delta Y(s), \ldots, -s\Delta Y(s), hs^{m-1}e^{-\theta s}, \ldots, he^{-\theta s}, he^{-\theta s}/s]^T,\\ \gamma=[a_n, a_{n-1}, \ldots, a_1, b_m, \ldots, b_1, b_0]^T.\end{cases} \tag{54}$$

one can obtain a weighted LS solution for parameter estimation,

$$\gamma=(\bar{\Phi}^T W\bar{\Phi})^{-1}\bar{\Phi}^T W\bar{\Psi} \tag{55}$$

where $\gamma=[a_n, a_{n-1}, \ldots, a_1, b_m, \ldots, b_1, b_0]^T$, $W=diag\{\rho_1, \ldots, \rho_M, \rho_1, \ldots, \rho_M\}$,

$$\bar{\Psi}=\begin{bmatrix}Re[\Psi]\\ Im[\Psi]\end{bmatrix},\quad \bar{\Phi}=\begin{bmatrix}Re[\Phi]\\ Im[\Phi]\end{bmatrix},$$

$$\Psi=[\psi(\alpha+j\omega_1), \psi(\alpha+j\omega_2), \ldots, \psi(\alpha+j\omega_M)]^T,\quad \Phi=[\phi(\alpha+j\omega_1), \phi(\alpha+j\omega_2), \ldots, \phi(\alpha+j\omega_M)]^T.$$

It can easily be verified that all the columns of $\bar{\Phi}$ are linear independent from each other, so that $(\bar{\Phi}^T W \bar{\Phi})^{-1}$ is guaranteed to be non-singular. Accordingly, there exists a unique solution for parameter estimation. For implementation, $M \in [10, 50]$ may be taken for a good trade-off between identification accuracy and computation efficiency.

Note that a prior knowledge of the process time delay (θ) is needed to derive the remaining parameters from (55). A one-dimensional search of θ within a possible range as can be observed from the step test may be performed in terms of the following fitting criterion to obtain the optimal parameter estimation,

$$Err = \frac{1}{N_s}\sum_{k=1}^{N_s}[y(kT_s) - \hat{y}(kT_s)]^2 < \varepsilon \tag{56}$$

where $y(kT_s)$ and $\hat{y}(kT_s)$ denotes the process and model outputs to the step change, respectively, and N_sT_s is the settling time.

It was clarified in [32] that consistent parameter estimation can be guaranteed by the above algorithm in combination with the time domain fitting criterion of (56).

From (54) it can be seen that only a single integral is needed in the above frequency response algorithm for the identification of a nth order model ($n \geq 2$). The single integral can be computed relatively independent of the time length of the step response. For model fitting, only M points of frequency response estimation are used in comparison with the previous time integral approach based on using multiple integrals on all the sampled transient response data. The computation effort is therefore between the two- or three-point fitting method and the time integral approach.

For illustration, the identification algorithms in Ref. [32] are used to identify the first-, second-, and fifth-order processes (G_1, G_2, and G_3) listed in Table 1 for comparison with the other categorized methods. For the fifth-order process (G_3), it is seen from Fig. 5 that desirable fitting accuracy is obtained by the three-point fitting method [19], the time integral approach [26], or the frequency response fitting method [32]. From the control results of ISE listed in Table 2, it is seen that the recent time integral approach [26] and frequency response fitting method [32] facilitate the IMC design obtaining better control performance of both the setpoint tracking and load disturbance rejection.

3.4. Robust identification under nonzero initial conditions and load disturbance

In practical applications, due to the presence of measurement noise or unexpected load disturbance, it is often difficult to know if the process to be identified has moved into a steady state that is suitable to perform a step test as required for application of the aforementioned identification methods. Besides, waiting for such a ‘steady state’ to have a step test can be quite troublesome for industrial processes with slow time constant or long time delay.

To cope with unknown initial conditions, Ahmed et al. [34] extended the time integral approach by taking the initial states of the process output and its derivatives as part of the parameters to be identified while assuming no load disturbance. The key idea was that sufficient LS based fitting conditions were constructed for model parameter estimation by performing multiple integrals to the time domain step response expression. For example, consider the identification of an FOPDT model, of which the output response to a step change has been shown in (11).

Denote multiple integrals for a time function of $f(t)$ as

$$\int_{[0,t]}^{(m)} f(t) = \int_0^t \int_0^{\tau_{m-1}} \cdots \int_0^{\tau_1} f(\tau_0)d\tau_0 d\tau_1 \cdots d\tau_{m-1}, \quad m \geq 2 \tag{57}$$

Doubly integrating both sides of (11) yields

$$\int_{[0,t]}^{(2)} y(t) = k_p h \left(\frac{t^2}{2} - \theta t + \frac{\theta^2}{2}\right) + \tau_p y(0)t - \tau_p \int_0^t y(\tau)d\tau \tag{58}$$

which can be reformulated into an LS form of (13), where

$$\begin{cases} \psi(t) = \int_{[0,t]}^{(2)} y(t), \\ \phi(t) = \left[-\int_0^t y(\tau)d\tau, ht^2/2, t, h/2\right]^T, \\ \gamma = [\tau_p, k_p, -hk_p\theta + \tau_p y(0), k_p\theta^2]^T. \end{cases} \tag{59}$$

Using the LS algorithm in (17) and (18), the FOPDT model parameters together with the initial output value can be explicitly estimated in terms of (59). To cope with measurement noise, an IV for the identification of a nth order process was suggested [34] as

$$z(t) = \left[-\int_0^t \hat{y}(\tau)d\tau, -\int_{[0,t]}^{(2)} \hat{y}(t), \cdots - \int_{[0,t]}^{(n)} \hat{y}(t), \frac{ht^n}{n!}, \frac{t^{n-1}}{(n-1)!}, \cdots, t, 1\right] \tag{60}$$

where $\hat{y}(t)$ denotes the model predicted output response.

Because redundant fitting conditions may be derived from such an algorithm using multiple integrals, it was suggested [34] to take the mean value for multiple estimations on the same model parameter. An alternative identification algorithm was proposed in [35] based on using a linear filter to construct the IV matrix and an iterative procedure to estimate the model parameters including the time delay.

For model identification subject to unknown initial process conditions and load disturbance, a modified time integral identification method was proposed by Wang et al. [36], where the integral limits were specially chosen such that the integral terms can be computed

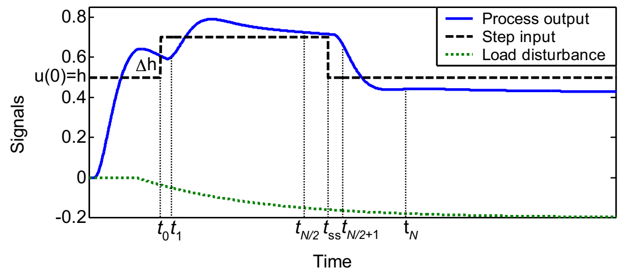

Fig. 8. A modified step test under nonzero initial process conditions and load disturbance.

independent of the unknown initial conditions. For illustration, consider the identification of an SOPDT model for a stable process which corresponds to the following step response equation,

$$a_2\ddot{y}(t)+a_1\dot{y}(t)+y(t)=b_1\dot{u}(t-\theta)+b_0u(t-\theta)+l \tag{61}$$

where l denotes a static disturbance or a bias regarding the process output in a steady state.

A modified double-integral computation to the derivatives of the output response is introduced as

$$\int_0^{\tau}\left[\int_{t-\tau_1}^{t+\tau_1}\dot{y}(\tau_0)d\tau_0\right]d\tau_1=\int_0^{\tau}[y(t+\tau_1)-y(t-\tau_1)]\,d\tau_1 \tag{62}$$

$$\int_0^{\tau}\left[\int_{t-\tau_1}^{t+\tau_1}\ddot{y}(\tau_0)\,d\tau_0\right]d\tau_1=\int_0^{\tau}[\dot{y}(t+\tau_1)-\dot{y}(t-\tau_1)]\,d\tau_1=y(t+\tau)-2y(t)+y(t-\tau) \tag{63}$$

where τ is a user-specified value and the constraint, $t-\tau<\theta\le t$, should be satisfied for all the sampled moments, $t_1<t_2<\cdots<t_N$, for computation. Note that $t_1-\tau<0$ is allowed if the output measurement before the step test can be obtained.

By integrating both sides of (61) in terms of the above double-integral computation, the resulting expression can be formulated into an LS form of (13), where

$$\begin{cases}\psi(t)=\int_0^{\tau}\left[\int_{t-\tau_1}^{t+\tau_1}y(\tau_0)\,d\tau_0\right]d\tau_1,\\ \phi(t)=\left[-\int_0^{\tau}\left[\int_{t-\tau_1}^{t+\tau_1}\ddot{y}(\tau_0)\,d\tau_0\right]d\tau_1,-\int_0^{\tau}\left[\int_{t-\tau_1}^{t+\tau_1}\dot{y}(\tau_0)\,d\tau_0\right]d\tau_1,-\frac{ht^2}{2},-t,\frac{h}{2}\right]^T,\\ \gamma=\left[a_2,a_1,b_0,b_1-b_0(\tau+\theta),b_0(\tau^2-2\tau\theta-\theta^2)+2b_1(\tau+\theta)+\frac{2l\tau^2}{h}\right]^T.\end{cases} \tag{64}$$

Besides, it follows from the steady state of (61) that

$$y(\infty)=b_0h+l \tag{65}$$

where $y(\infty)$ is estimated from the steady-state output value of the step response.

Hence, the model parameters can be derived by solving (64) and (65). It can be seen from (65) that only static load disturbance is allowed for the use of such identification algorithm.

To cope with nonzero or unknown initial conditions and/or non-static load disturbance, a robust identification method was proposed by Liu and Gao [37] based on a modified step test. Such an identification test is illustrated in Fig. 8, where t_0 is the time for adding a step change, t_{ss} the time for removing the step change, and $t_1<t_2<\cdots<t_{N/2}<t_{N/2+1}<\cdots<t_N$ the sampled moments. Both the transient response data from adding and subsequently removing the step change are collected for model identification. According to the modified step test where

$$u(t)=\begin{cases}h, & 0\le t\le\theta;\\ h+\Delta h, & \theta<t\le t_{ss}+\theta;\\ h, & t>t_{ss}+\theta.\end{cases} \tag{66}$$

it can be derived that

$$\int_0^{t}u(\tau)\,d\tau=\begin{cases}(h+\Delta h)t-\Delta h\theta, & \theta<t\le t_{ss}+\theta;\\ \Delta ht_{ss}+ht, & t>t_{ss}+\theta.\end{cases} \tag{67}$$

$$\int_0^{t}\dot{u}(\tau)\,d\tau=\begin{cases}\Delta h, & \theta<t\le t_{ss}+\theta;\\ 0, & t>t_{ss}+\theta.\end{cases} \tag{68}$$

$$\int_{[0,t]}^{(2)} u(t) = \begin{cases} \dfrac{h+\Delta h}{2}t^2 - \Delta ht\theta + \dfrac{\Delta h}{2}\theta^2, & \theta < t \le t_{ss}+\theta; \\ \dfrac{h}{2}t^2 + \Delta ht_{ss}t - \dfrac{\Delta h}{2}t_{ss}^2 - \Delta ht_{ss}\theta, & t > t_{ss}+\theta. \end{cases} \tag{69}$$

$$\int_{[0,t]}^{(2)} \dot{u}(t) = \begin{cases} \Delta h(t-\theta), & \theta < t \le t_{ss}+\theta; \\ \Delta ht_{ss}, & t > t_{ss}+\theta. \end{cases} \tag{70}$$

For identifying an FOPDT model, by doubly integrating both sides of (11) and rearranging the resulting equation using (69) and (70), one can formulate an LS fitting in the form of (13), where

$$\begin{cases} \psi(t) = \displaystyle\int_{[0,t]}^{(2)} y(t), \\ \phi(t) = \left[-\displaystyle\int_0^t y(\tau)\,d\tau, F_0(t), F_1(t), F_2(t), 1, t, t^2, \cdots, t^q\right]^T, \\ \gamma = \left[\tau_p, k_p, k_p\theta, k_p\theta^2, \eta_0, \eta_1, \eta_2, \cdots, \eta_q\right]^T. \end{cases} \tag{71}$$

$$F_0(t) = \begin{cases} \dfrac{h+\Delta h}{2}t^2, & \theta < t \le t_{ss}+\theta; \\ \dfrac{h}{2}t^2 + \Delta ht_{ss}t - \dfrac{\Delta h}{2}t_{ss}^2, & t > t_{ss}+\theta. \end{cases} \tag{72}$$

$$F_1(t) = \begin{cases} -\Delta ht, & \theta < t \le t_{ss}+\theta; \\ -\Delta ht_{ss}, & t > t_{ss}+\theta. \end{cases} \tag{73}$$

$$F_2(t) = \begin{cases} \dfrac{\Delta h}{2}, & \theta < t \le t_{ss}+\theta; \\ 0, & t > t_{ss}+\theta. \end{cases} \tag{74}$$

where $\eta_i\,(i = 0, 1, 2, \ldots, q)$ in the parameter vector γ are used for the Maclaurin series approximation of the unforced output response resulting from nonzero initial process conditions and load disturbance, according to the superposition principle for analysis of system response. For consistent estimation against measurement noise, an IV was suggested [37] as $z_i = [1/t_i^4, 1/t_i^3, 1/t_i^2, 1/t_i, 1, t_i, t_i^2, \cdots, t_i^q]^T$.

An important merit of such an LS algorithm is that the excitation sequence and the corresponding output sequence consisting of two segments of the test data, are no longer correlated to the time sequence or any other time-related sequences like the transient response sequence resulting from nonzero initial process conditions and/or load disturbance. It has therefore been clarified [37–39] that a conventional step test is not capable of consistent parameter estimation under unsteady initial process conditions and/or non-static load disturbance.

In the presence of a deterministic load disturbance, which is often confronted in operating industrial control systems that are frequently or periodically initiated by a step change to the setpoint, the resulting step response may be viewed as the pure process response plus the load disturbance response according to the superposition principle. For instance, a water pump in an air conditioning system turns out different step responses under different loads. Such load disturbance is herein named an inherent-type disturbance. Modeling both the pure pump response to the setpoint without load and the disturbance response of load can facilitate the control design for the pump operation under a variety of load levels. A piecewise model identification method was proposed by Liu et al. [40] to simultaneously obtain low-order models of both the process and the load disturbance from a single step test. A guideline for partitioning the step response data was given for developing such an identification method. Based on the guideline, the previous time integral approach can be extended to directly use the raw step response data for model identification.

3.5. Model identification from closed-loop step test

There are in general two motivations for using a closed-loop step test for model identification, one is for online tuning a closed-loop control system prescribed for a stable process, and another for identifying the fundamental dynamic response characteristics (usually in the low frequency range) of an integrating or unstable process subject to rigorous operation conditions. A practical restriction is that the closed-loop controller needs to be specified in advance to maintain the closed-loop stability under such a test. To perform a closed-loop step test, the step change may be added to either the setpoint or the process input (corresponding to the controller output) in terms of the closed-loop structure for a stable process, but should be confined to only the setpoint for an integrating or unstable process, owing to the requirement on the closed-loop stability [41].

For closed-loop step identification, when the process input is measured as well as the process output, all the aforementioned step identification methods can be used to identify the process model, e.g. the time integral approach developed by Li et al. [42] for the identification of multiple-input-multiple-output (MIMO) processes based on closed-loop step tests.

When only the process output is measured, i.e. only the data from the setpoint change and the output response can be used for model identification, a two-step identification procedure was generally adopted in the existing references: The first step is the identification of the overall closed-loop system transfer function or frequency response. The aforementioned model identification methods and frequency response estimation methods can be used for this purpose. The second step is the computation of the process model or frequency response from the closed-loop system transfer function, based on a prior knowledge of the controller setting. With a known or specified model structure for the process, sufficient fitting conditions can be established to drive the model parameters. For example, consider the unity feedback control system shown in Fig. 2, where the closed-loop transfer function can be derived as

$$T(s) = \frac{G(s)C(s)}{1+G(s)C(s)} \tag{75}$$

With a known form of $C(s)$, the process transfer function can be inversely derived from (75) as

$$G(s) = \frac{T(s)}{C(s)[1-T(s)]} \tag{76}$$

To avoid model complexity from the above derivation, it was suggested to use a simple P-type controller to derive a low-order model such as FOPDT, by using the time domain response data to establish the fitting conditions in terms of the characteristic equation of the closed-loop transfer function [43–46]. In contrast, it was proposed to first estimate the process frequency response from the above expression, and then establish model fitting conditions based on the frequency response estimation [41,47–52]. For illustration, assume that the controller is a conventional PID controller,

$$C(s) = k_C + \frac{1}{\tau_I s} + \frac{\tau_D s}{\tau_F s + 1} \tag{77}$$

where k_C denotes the controller gain, τ_I the integral constant, τ_D the derivative constant, and τ_F a filter constant that is usually taken as $\tau_F = (0.01–0.1)\tau_D$ for implementation.

It can be derived in terms of the definition in (38) that

$$C^{(1)}(s) = -\frac{1}{\tau_I s^2} + \frac{\tau_D}{(\tau_F s + 1)^2} \tag{78}$$

$$C^{(2)}(s) = \frac{2}{\tau_I s^3} - \frac{2\tau_D \tau_F}{(\tau_F s + 1)^3} \tag{79}$$

Using the frequency response estimation method presented in Section 3.3, the closed-loop system frequency response can be estimated by

$$T(\alpha + j\omega) = \frac{\alpha + j\omega}{h} \Delta Y(\alpha + j\omega), \quad \alpha > 0 \tag{80}$$

and its first and second derivatives with respect to $s = \alpha + j\omega$ can be computed by

$$T^{(1)}(s) = \frac{1}{h}\int_0^{t_N} (1 - st)\Delta y(t) e^{-st}\, dt \tag{81}$$

$$T^{(2)}(s) = \frac{1}{h}\int_0^{t_N} t(st-2)\Delta y(t) e^{-st}\, dt \tag{82}$$

where t_N can be chosen in terms of the numerical constraints shown in (41) and (42).

Correspondingly, the first and second derivatives of $G(s)$ can be derived from (76) as

$$G^{(1)} = \frac{T^{(1)}C + C^{(1)}T(T-1)}{C^2(1-T)^2} \tag{83}$$

$$G^{(2)} = \frac{CT^{(2)} + 2C^{(1)}T^{(1)}T + C^{(2)}T(T-1)}{C^2(1-T)^2} - \frac{2[CT^{(1)} + C^{(1)}T(T-1)][C^{(1)}(1-T) - CT^{(1)}]}{C^3(1-T)^3} \tag{84}$$

Hence, the model fitting algorithms based on frequency response estimation as presented in Section 3.3 can be adopted afterward to derive a model such as FOPDT or SOPDT, which are omitted for brevity.

4. Relay identification methods

Since the pioneering work occurred in the 1980s [1–4], model identification and controller autotuning from relay feedback have attracted gradually increasing attentions in the process control community. Compared to a step test, the fundamental dynamic response characteristics of the process can be substantially observed from the sustained oscillations under relay feedback, especially in the presence of measurement noise. Moreover, a relay feedback test will not cause the process to drift too far away from its setpoint, as well as the traditional 'Ziegler–Nichols' autouning approach based on using the closed-loop oscillation condition of ultimate gain [53]. This is very necessary in many practical applications, in particular for a piecewise identification of a highly nonlinear process that is subject to rigorous operation conditions [4].

Generally, there are two types of relay tests – unbiased (symmetrical) and biased (asymmetrical). A biased relay function is expressed by

$$u(t) = \begin{cases} u_+ & \text{for } \{e(t) > \varepsilon_+\} \text{ or } \{e(t) \geq \varepsilon_- \text{ and } u(t_-) = u_+\} \\ u_- & \text{for } \{e(t) < \varepsilon_-\} \text{ or } \{e(t) \leq \varepsilon_+ \text{ and } u(t_-) = u_-\} \end{cases} \tag{85}$$

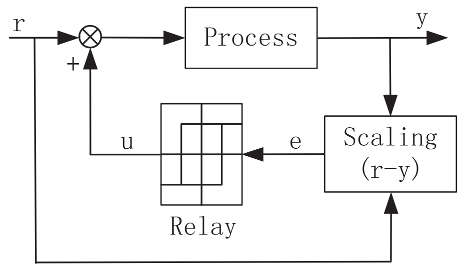

Fig. 9. Schematic of an online relay test.

where $u_+ = \Delta\mu + \mu_0$ and $u_- = \Delta\mu - \mu_0$ denote, respectively, the positive and negative relay magnitudes; ε_+ and ε_- denote, respectively, the positive and negative switch hystereses. The initial output of the relay is assumed as u_- for zero input, as commonly set in industrial applications.

Note that letting $\Delta\mu = 0$ and $|\varepsilon_+| = |\varepsilon_-|$ leads to an unbiased relay function. For the use of $\Delta\mu = 0$ and $\varepsilon_+ = \varepsilon_- = 0$, it is customarily named as an 'ideal' relay test in the literature.

A typical configuration for a relay feedback test has been shown in Fig. 3, where *r* denotes the setpoint, *y* the process output, and *u* the relay output. To capture the process dynamic response characteristics around different setpoints in operation, an online relay test may be constructed as shown in Fig. 9, where the scaling unit (*r*–*y*) is set to normalize *r* and *y* for computing the output error, *e*, under nonzero initial process conditions.

Because a relay feedback test is under a closed-loop control structure, the closed-loop stability must be taken into account for implementation. For the identification of stable and integrating processes, it was clarified by deriving the relay response expressions [54–59] that sustained oscillations leading to the limit cycle can surely be formed for such a process. For the identification of unstable processes, some limiting conditions to forming steady oscillations under an unbiased relay test were reported in Refs. [60–65], and for the use of a biased relay test, a limiting condition to forming steady oscillation for a first-order unstable process was revealed in [58]. To ensure the relay feedback stability, two relay identification methods based on using a P-, or PD-type controller for closed-loop stabilization were developed in [62,66].

Given the process output response in the limit cycle under a relay test, a suitable model structure for fitting cannot be intuitively determined as under a step test, since different model structures had been traditionally defined based on an open-loop step test in practical applications. To solve the problem, a curvature factor was defined by Luyben [67] to distinguish different relay response shapes for the choice of a suitable FOPDT model structure to identify a stable or integrating process. By performing an 'ideal' relay test to a variety of stable processes with different damping ratios or dead-time-vs-time-constant ratios, a few graphical guidelines were proposed by Panda and Yu [68] for choosing a suitable FOPDT or SOPDT model to fit the relay response. Furthermore, some guidelines for choosing a suitable SOPDT model under an unbiased or biased relay test was given by Liu et al. [57] for fitting various response types of stable processes. For the identification of an integrating or unstable process, it was suggested in [58] that the use of an SOPDT model shown in (3) or (4) could obtain obviously better fitting effect compared with an FOPDT model, but a higher-order model might not facilitate improving the fitting accuracy with respect to the fundamental dynamic response characteristics.

Looking into the technical routes for model identification in the existing references, we categorize the developed relay identification methods into three groups, one is the DF method, another is the curve fitting approach, and the third is the use of frequency response estimation for model fitting. These categorized identification methods are reviewed in the following subsections, respectively, along with two subsections specifically for summarizing robust relay identification methods against measurement noise and/or load disturbance, and those identification methods for MIMO processes.

4.1. Describing function method

The DF method was among the pioneering work for relay identification, and has been continuously developed up to the present [13,14]. The key idea behind this approach is the use of the ultimate amplitude and phase conditions for steady oscillations under a relay test to establish fitting conditions for model identification. Denote by $N(A)$ the DF of a relay module shown in Fig. 3, where A is the input of the relay module, i.e. the output deviation from the setpoint. In the limit cycle under an unbiased relay test, with the measured oscillation amplitude A_+ and period P_u of the output, the following ultimate amplitude and phase conditions were constructed [1–3] by ignoring the influence from the output harmonics in the steady oscillations (i.e. the harmonic terms in the Fourier expansion of the limit cycle),

$$\left|N(A_+)G(j\omega_u)\right| = 1 \tag{86}$$

$$\angle N(A_+) + \angle G(j\omega_u) = -\pi \tag{87}$$

where $\omega_u = 2\pi/P_u$, and

$$N(A_+) = \frac{4\mu_0}{\pi A_+} e^{-j\arcsin(\varepsilon_+/A_+)} \tag{88}$$

Note that μ_0 and ε_+ have been defined by (85).

For the use of an 'ideal' relay test, it follows that

$$\left|G(j\omega_u)\right| = \frac{\pi A_+}{4\mu_0} \tag{89}$$

$$\angle G(j\omega_u) = -\pi \tag{90}$$

which have been adopted for model identification and frequency response estimation in many existing references, e.g. [4,6,12–14,69,70].

Table 3
Illustrative comparison between the categorized relay identification methods.

Process	NSR (%)	Describing function method	Curve fitting approach	Frequency response fitting
$G_4 = \frac{1}{10s+1}e^{-2s}$	0	$\frac{1}{8.118s+1}e^{-2s}$ (Ref. [4])	$\frac{1.0001}{9.9954s+1}e^{-2s}$ (Ref. [54])	$\frac{1.0001}{10.001s+1}e^{-2.005s}$ (Ref. [56])
$G_5 = \frac{1}{s-1}e^{-0.4s}$	0	$\frac{0.928}{0.757s-1}e^{-0.392s}$ (Ref. [74])	$\frac{1.0001}{0.9976s-1}e^{-0.4s}$ (Ref. [56])	$\frac{1.0001}{1.0009s-1}e^{-0.4038s}$ (Ref. [56])
$G_6 = \frac{1}{(s+1)(10s+1)}e^{-2s}$	0	$\frac{0.853e^{-2s}}{(1.15s+1)(7.416s+1)}$ (Ref. [69])	$\frac{1.0000e^{-1.9991s}}{(1.0073s+1)(9.9921s+1)}$ (Ref. [57])	$\frac{1.05}{(1.271s+1)(9.766s+1)}e^{-1.814s}$ (Ref. [90])
	5	$\frac{0.746e^{-2.2s}}{(1.852s+1)(8.635s+1)}$ (Ref. [69])	$\frac{0.9943e^{-1.9329s}}{(1.098s+1)(9.8346s+1)}$ (Ref. [57])	$\frac{1.06}{(0.92s+1)(10.73s+1)}e^{-2.04s}$ (Ref. [90])
$G_7 = \frac{1}{s(20s+1)}e^{-10s}$	0	–	$\frac{1.0000}{s(20.0047s+1)}e^{-10.0097s}$ (Ref. [81])	$\frac{0.9983}{s(19.9443s+1)}e^{-10.027s}$ (Ref. [58])
$G_8 = \frac{-s+1}{(s+1)^5}e^{-s}$	0	$\frac{1.0}{(1.48s+1)^2}e^{-4.56s}$ (Ref. [72])	$\frac{1.0}{(1.4767s+1)^2}e^{-4.125s}$ (Ref. [78])	$\frac{0.9992}{(1.4781s+1)^2}e^{-4.037s}$ (Ref. [59])
$G_9 = \frac{e^{-0.5s}}{(5s-1)(2s+1)(0.5s+1)}$	0	$\frac{1.0}{5.77s-1}e^{-11.97s}$ (Ref. [74])	$\frac{1.001e^{-0.939s}}{10.354s^2+2.932s-1}$ (Ref. [80])	$\frac{1.0002e^{-0.94s}}{(4.9997s-1)(2.0696s+1)}$ (Ref. [58])
Computation effort		Low	Moderate or relatively high	Moderate or relatively high

For illustration, consider the identification of an FOPDT model for a stable process, by substituting (1) into (89) and (90), the time constant and static gain can be derived using the time delay directly observed from the relay test [69], respectively, as

$$\tau_p = \frac{1}{\omega_u}\tan(\pi - \theta\omega_u) \tag{91}$$

$$k_p = \frac{\pi A_+}{4\mu_0}\sqrt{\tau_p^2\omega_u^2 + 1} \tag{92}$$

Similarly, an FOPDT model can be obtained for an unstable process except for that $\tau_p = \tan(\theta\omega_u)/\omega_u$ should be derived instead of (91) according to the phase condition in (90).

Since the above two fitting conditions are not sufficient for deriving a higher-order model with more parameters, additional relay tests and fitting conditions were suggested in [69,71–74] for parameter estimation. It should be noted that although the above ultimate amplitude and phase conditions are simple for model identification and online controller tuning, the identification accuracy is indeed not high due to the use of approximation for the relay function in the DF analysis and the overlook on the harmonic components of the limit cycle, especially for identifying higher-order or time delay processes of which the relay responses are apparently different from the sinusoidal shape. By compensating for the phase lag caused by the relay module, a modified DF identification method was proposed in [75] which can obtain significantly improved accuracy for higher-order or large time delay processes.

For illustration, the DF methods developed in Refs. [4,69,72,74] are used to identify the first-, second-, and high-order processes (G_4–G_9) listed in Table 3, for which the relay tests are performed the same as those detailed in Refs. [4,54,56,58,69,80]. Note that for the presence of measurement noise with NSR = 5%, the averaged limit cycle data of 20 steady oscillation periods are used to obtain the identification results for G_6 as listed in Table 3. For the high-order stable and unstable processes (G_8 and G_9), the Nyquist fitting effects of the identified models are shown in Figs. 10 and 11, respectively. To illustrate the achievable control performance for the high-order unstable process (G_9), the IMC-based two-degree-of-freedom (2DOF) control method [8] is adopted based on the identified FOPDT and SOPDT models as listed in Table 3. By adding a step change of the setpoint and assuming a step-type load disturbance with a magnitude of −0.1 that occurs at $t = 600(s)$ from the process input side, the control results are shown in Fig. 12, where the controller parameters are taken as $\lambda_c = \lambda_f = 8$ for the setpoint tracking and load disturbance rejection, respectively.

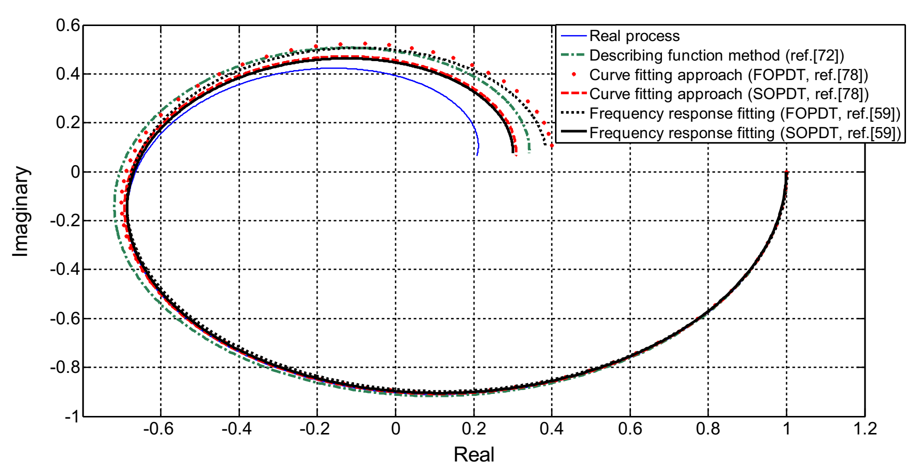

Fig. 10. Nyquist fitting of the low-order models identified by the categorized relay methods for a high-order stable process.

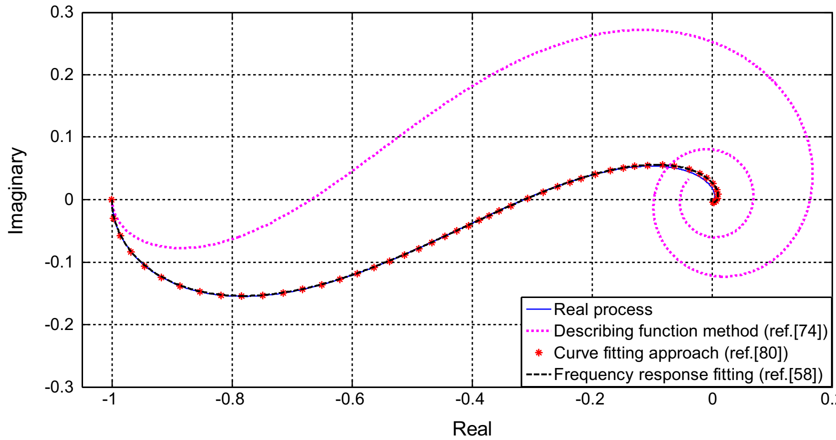

Fig. 11. Nyquist fitting of the low-order models identified by the categorized relay methods for a high-order unstable process.

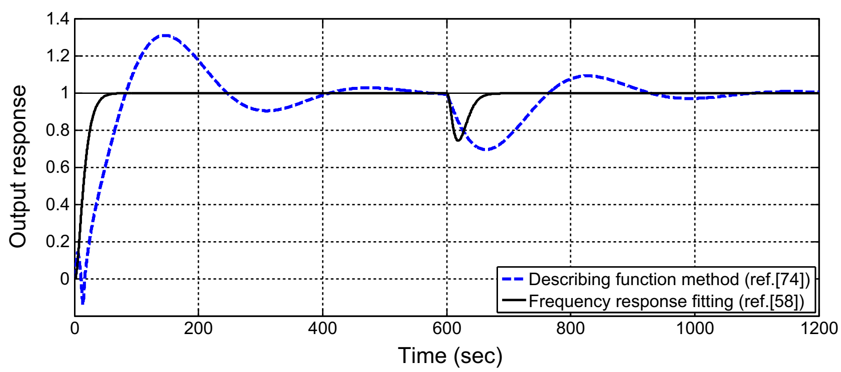

Fig. 12. Illustrative comparison of the categorized relay identification methods for the IMC-based control design for a high-order unstable process.

4.2. Curve fitting approach

In the steady oscillations under a relay test, there are observable time points in the limit cycle curve, including the oscillation amplitudes, the moments to reach the oscillation amplitudes, the relay switching moments and the corresponding output response points, and the half oscillation periods. To make use of these observable data for model identification, analytical expressions of the relay response have been derived in terms of different low-order models to construct the corresponding fitting conditions for parameter estimation in the literature [54–59,76,77]. For example, consider the identification of an FOPDT model for an overdamped stable process based on a biased relay test (using the relay function in (85)) as shown in Fig. 13. It can be derived that the resulting limit cycle of the output response is characterized [56] by

$$y_{+}(t) = k_{\mathrm{p}}(\Delta\mu + \mu_0) - 2k_{\mathrm{p}}\mu_0 E e^{-(t/\tau_{\mathrm{p}})}, \quad t \in [0, P_{+}] \tag{93}$$

$$y_{-}(t) = k_{\mathrm{p}}(\Delta\mu - \mu_0) - 2k_{\mathrm{p}}\mu_0 F e^{-(t/\tau_{\mathrm{p}})}, \quad t \in [0, P_{-}] \tag{94}$$

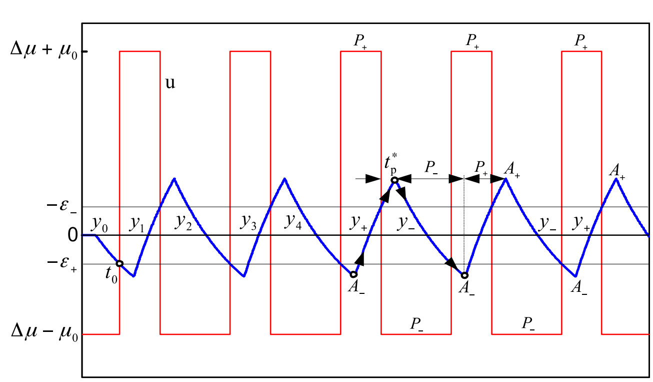

Fig. 13. The limit cycle for identifying an FOPDT model.

where $y_+(t)$ is the monotonically ascending part for $t \in [0, P_+]$, $y_-(t)$ is the monotonically descending part for $t \in [0, P_-]$ that corresponds to $t \in [P_+, P_u]$ in the limit cycle, $P_u = P_+ + P_-$ is the oscillation period, and

$$E = \frac{1 - e^{-(p_-/\tau_p)}}{1 - e^{-(p_u/\tau_p)}} \tag{95}$$

$$F = -\frac{1 - e^{-(p_+/\tau_p)}}{1 - e^{-(p_u/\tau_p)}} \tag{96}$$

If an unbiased relay test is used, substituting $\Delta\mu = 0$ and $P_+ = P_- = P_u/2$ into (93)–(96) yields the corresponding relay response expression,

$$y_+(t) = -y_-(t) = k_p\mu_0 - 2k_p\mu_0 E e^{-(t/\tau_p)} \tag{97}$$

where $y_+(t)$ is for $t \in [0, P_u/2]$, $y_-(t)$ is for $t \in (P_u/2, P_u]$, and

$$E = \frac{1}{1 + e^{-(p_u/2\tau_p)}} \tag{98}$$

Note that, in the limit cycle the process output is a periodic function with respect to the oscillation angular frequency, $\omega_u = 2\pi/P_u$. Using a time shift, one can view it as a periodic signal from the very beginning, so its Fourier transform can be derived as

$$Y(j\omega_u) = \lim_{N\to\infty} N \int_0^{P_u} y_{os}(t) e^{-j\omega_u t}\,dt = \lim_{N\to\infty} N \int_{t_{os}}^{t_{os}+P_u} y(t) e^{-j\omega_u t}\,dt \tag{99}$$

where $y_{os}(t) = y(t)$ for $t \in [t_{os}, \infty)$ and t_{os} can be taken as any relay switching point in the steady oscillations, such that the influence from the initial process response to the above periodic integral can be excluded.

Thereby, the process frequency response at ω_u can be obtained as

$$G(j\omega_u) = \frac{Y(j\omega_u)}{U(j\omega_u)} = \frac{\int_{t_{os}}^{t_{os}+P_u} y(t) e^{-j\omega_u t}\,\mathrm{d}t}{\int_{t_{os}}^{t_{os}+P_u} u(t) e^{-j\omega_u t}\,\mathrm{d}t} = A_u e^{j\varphi_u} \tag{100}$$

Note that the numerical integral in (100) may be computed using the trapezoidal rule or the FFT for $Y(j\omega_u)$ and $U(j\omega_u)$.

When a biased relay test is used, the process static gain can be derived from (100) as

$$k_p = G(0) = \frac{\int_{t_{os}}^{t_{os}+P_u} y(t)\,dt}{\int_{t_{os}}^{t_{os}+P_u} u(t)\,dt} \tag{101}$$

It can be seen from (93) and (94) that the process time delay can directly be read as the time taken to reach the output response peak from the initial relay switching point in a half period of the relay, which is denoted by t_p^* shown in Fig. 13. Correspondingly, it follows from (93) and (94) that

$$y_+(P_+) = k_p(\Delta\mu + \mu_0) - 2k_p\mu_0 E e^{-(P_+/\tau_p)} = A_+ \tag{102}$$

$$y_-(P_-) = k_p(\Delta\mu - \mu_0) - 2k_p\mu_0 F e^{-(P_-/\tau_p)} = A_- \tag{103}$$

$$y_+(P_+ - \theta) = k_p(\Delta\mu + \mu_0) - 2k_p\mu_0 E e^{-((P_+ - \theta)/\tau_p)} = -\varepsilon_- \tag{104}$$

$$y_-(P_- - \theta) = k_p(\Delta\mu - \mu_0) - 2k_p\mu_0 F e^{-((P_- - \theta)/\tau_p)} = -\varepsilon_+ \tag{105}$$

Substituting (102) into (104) yields

$$\tau_p = \frac{\theta}{\ln((k_p(\Delta\mu + \mu_0) + \varepsilon_-)/(k_p(\Delta\mu + \mu_0) - A_+))} \tag{106}$$

Alternatively, substituting (103) into (105) yields

$$\tau_p = \frac{\theta}{\ln((k_p(\Delta\mu - \mu_0) + \varepsilon_+)/(k_p(\Delta\mu - \mu_0) - A_-))} \tag{107}$$

Therefore, the process time constant can be derived from (106) or (107). It is preferred in practice to use (106) for better accuracy in case of model mismatch, owing to that the positive fitting part in a half period of the limit cycle occupies a larger percentage compared to the negative fitting part in the other half period, and vice versa.

When an unbiased relay test is used, the process static gain cannot be derived from (101) since a periodic integral of the relay or the process output response becomes zero. Nevertheless, the time delay parameter can also be read as t_p^* shown in Fig. 13, according to the relay response expression in (97). Solving (102) and (104) together to eliminate k_p yields

$$\varepsilon_+(1 - e^{-(p_u/(2\tau_p))}) = A_+(1 + e^{-(p_u/(2\tau_p))} - 2e^{-((p_u - 2\theta)/(2\tau_p))}) \tag{108}$$

Note that the left-hand side of (108) is monotonically decreasing with respect to τ_p, together with a practical constraint of $0 < \tau_p < P_u$. Only finite solutions of τ_p therefore exist for (108), which can be derived using any iterative algorithm such as the Newton–Raphson method. The initial estimation for iteration may be taken as $\hat{\tau}_p = P_u/2 - \theta$, in consideration of that the influence from the process time constant to the relay response corresponds to this time interval.

The process static gain can then be derived from (102) as

$$k_p = \frac{A_+(1+e^{-(p_u/(2\tau_p))})}{\mu_0(1-e^{-(p_u/(2\tau_p))})} \tag{109}$$

In case multiple solutions of τ_p and k_p are obtained from the above computation, a suitable solution pair of τ_p and k_p can be determined by comparing the relay response of the resulting model with that of the process, or using the ultimate oscillation conditions, $|N(A_+)\widehat{G}(j\omega_u)| \to 1$ and $\angle N(A_+) + \angle\widehat{G}(j\omega_u) \to -\pi$, where $\widehat{G}(j\omega_u)$ is the FOPDT model response at the oscillation frequency and $N(A_+) = 4\mu_0 e^{-j\arcsin(\varepsilon_+/A_+)}/(\pi A_+)$ as introduced in the aforementioned DF approach.

Note that with the process time delay read directly from the limit cycle, the process static gain and time constant can also be derived from the fitting conditions established at the oscillation frequency, that is, submitting the FOPDT model in (1) into (100) to match the estimated magnitude and phase values yields

$$\tau_p = \frac{1}{\omega_u} tan(-\varphi_u - \theta\omega_u), \quad \varphi_u \in (-\pi, -\pi/2) \tag{110}$$

$$k_p = A_u\sqrt{\tau_p^2\omega_u^2 + 1} \tag{111}$$

It can be seen from (110) that $\varphi_u \in (-\pi, -\pi/2)$ is used, rather than $\varphi_u = -\pi$ that had been traditionally used in the DF approach, thus facilitating the fitting accuracy at the oscillation frequency of the process response. Note that the phase condition, $\varphi_u = -\pi + \arctan\omega_u\tau_p - \theta\omega_u$, should be used to derive an FOPDT model for an unstable process, of which the identification formulae can be found in [8,56,58].

In contrast, a so-called 'A-locus' method [78] was proposed to fit the limit cycle under a biased relay test for model identification, based on measuring the positive and negative oscillation amplitudes, the moments to reach these amplitudes, and the oscillation periods. Nonlinear fitting equations were constructed for deriving the model parameters, which required numerical computation to obtain the solutions, therefore subject to computational sensitivity in the presence of measurement noise.

By deriving the state-space expression of the relay response under an 'ideal' relay test, a few low-order model identification algorithms were developed in [79,80] based on measuring the oscillation amplitudes, the moments to reach the amplitudes, and the oscillation periods, together with computing the area of the output response over a half period and/or the derivatives of the output response. Nonlinear fitting equations with respect to the limit cycle were also established to derive the model parameters. To overcome the above deficiency in the presence of measurement noise, a Fourier series expansion approach was suggested to filter the measured data for computation.

For illustration, the identification algorithms in Refs. [54,56,57,78,80,81] are used to identify the first-, second-, and high-order processes (G_4–G_9) listed in Table 3 for comparison with the other categorized methods, along with two comparisons on the Nyquist fitting of G_8 and G_9 as shown in Figs. 10 and 11, respectively.

4.3. Frequency response estimation for model fitting

Since the aforementioned DF method can only give a rough estimation of the process response at the oscillation frequency under a relay test, improved frequency response estimation methods were subsequently developed for model fitting and online controller tuning. By adding a parasitic relay loop into the conventional relay structure for test, multiple oscillation periods were obtained [82] for estimating the process response at different oscillation frequencies. The use of multiple relay modules in parallel was explored in [83,84] for identifying multiple frequency response points in a specified range. Based on a single relay test, a few algorithms were developed in [85–87] by letting $s = j\omega$ for estimating multiple frequency response points in the low frequency range, in particular for the range from zero frequency to the relay oscillation frequency. By comparison, two alternative algorithms were proposed in [59,88] by letting $s = \alpha + j\omega$ where α is regarded as a damping factor for numerically computing the Laplace transform with a finite time length. It was demonstrated in [59] that the use of $s = \alpha + j\omega$ can guarantee unbiased frequency response estimation from zero frequency to a higher frequency (but smaller than the half sampling frequency), and therefore, can be used to derive the process static gain from a single run of the relay test, independent of the relay type being unbiased or biased.

For illustration, consider that an unbiased relay test is performed. In the limit cycle the process output is a periodic function with respect to the oscillation angular frequency, $\omega_u = 2\pi/P_u$. Using an idea of time shift, it has been shown by the expression in (100) that the process frequency response at ω_u can be computed.

To estimate the process frequency response other than at ω_u, one can decompose the relay response as

$$Y(s) = \int_0^{t_{os1}} y(t)e^{-st}\,dt + \int_{t_{os1}}^{\infty} y(t)e^{-st}dt \tag{112}$$

where t_{os1} denotes the time to reach the steady oscillations under a relay test. For implementation, t_{os1} may be chosen as any time instant in the steady oscillations, e.g. a small multiple of P_u for convenience.

In view of that $y(t)$ becomes a periodic signal after t_{os1}, the second integral term in (112) can be derived as

$$\begin{aligned}\int_{t_{os1}}^{\infty} y(t)e^{-st}\,dt &= \int_{t_{os1}}^{t_{os1}+P_u} y(t)e^{-st}\,dt + \int_{t_{os1}+P_u}^{t_{os1}+2P_u} y(t)e^{-st}dt + \cdots \\ &= (1 + e^{-P_u s} + e^{-2P_u s} + \cdots)\int_{t_{os1}}^{t_{os1}+P_u} y(t)e^{-st}dt \\ &= \lim_{n\to\infty}\frac{1-e^{-nP_u s}}{1-e^{-P_u s}}\int_{t_{os1}}^{t_{os1}+P_u} y(t)e^{-st}dt\end{aligned} \tag{113}$$

Note that if Re(s) > 0, there exists $e^{-nP_u s} \to 0$ for $n \to \infty$. One can thus obtain

$$Y(s) = \int_0^{t_{os1}} y(t)e^{-st}\,dt + \frac{1}{1-e^{-P_u s}} \int_{t_{os1}}^{t_{os1}+P_u} y(t)e^{-st}\,dt \tag{114}$$

Similarly, it follows for Re(s) > 0 that

$$U(s) = \int_0^{t_{os1}} u(t)e^{-st}\,dt + \frac{1}{1-e^{-P_u s}} \int_{t_{os1}}^{t_{os1}+P_u} u(t)e^{-st}\,dt \tag{115}$$

The process frequency response transfer function for Re(s) > 0 can therefore be derived as

$$G(s) = \frac{Y(s)}{U(s)} = \frac{(1-e^{-P_u s})\int_0^{t_{os1}} y(t)e^{-st}\,dt + \int_{t_{os1}}^{t_{os1}+P_u} y(t)e^{-st}\,dt}{(1-e^{-P_u s})\int_0^{t_{os1}} u(t)e^{-st}\,dt + \int_{t_{os1}}^{t_{os1}+P_u} u(t)e^{-st}\,dt} \tag{116}$$

Substituting $s = \alpha + j\omega$ into (116) yields

$$G(\alpha+j\omega) = A_\alpha e^{j\varphi_\alpha} = \frac{(1-e^{-P_u(\alpha+j\omega)})\int_0^{t_{os1}} [y(t)e^{-\alpha t}]e^{-j\omega t}\,dt + \int_{t_{os1}}^{t_{os1}+P_u} [y(t)e^{-\alpha t}]e^{-j\omega t}\,dt}{(1-e^{-P_u(\alpha+j\omega)})\int_0^{t_{os1}} [u(t)e^{-\alpha t}]e^{-j\omega t}\,dt + \int_{t_{os1}}^{t_{os1}+P_u} [u(t)e^{-\alpha t}]e^{-j\omega t}\,dt} \tag{117}$$

where $\alpha \in \Re_+$ may be viewed as a damping factor for numerically computing the Laplace transform.

Note that $G(j\omega + \alpha) \to 0/0$ as $\alpha \to \infty$. A guideline for choosing α was therefore suggested [59] as

$$\begin{cases} \alpha > \delta; \\ \min\{\left|u(t_{os1}+P_u)e^{-\alpha(t_{os1}+P_u)}T_s\right|, \left|y(t_{os1}+P_u)e^{-\alpha(t_{os1}+P_u)}T_s\right|\} > \delta. \end{cases} \tag{118}$$

where δ denotes the computational precision that may be practically taken smaller than $10^{-6} \times \min\{\left|u(t_{os1}+P_u)T_s\right|, \left|y(t_{os1}+P_u)T_s\right|\}$, and T_s is the sampling period for computing the numerical integrals in (117).

Accordingly, one can compute multiple frequency response points of $G(\alpha + j\omega_k)$ from (117) for $\omega_k = k\omega_u/M_0$ ($k = 0, 1, 2, \ldots, N_0 - 1$), where $N_0 = M_0 P_u/T_s$ and M_0 should be taken as an even integer for efficient computation of the FFT series.

Using the inverse Laplace transform of $g(t)e^{-\alpha t} = L^{-1}[G(\alpha + j\omega)]$, one obtains

$$G(0) = \text{FFT}\{\text{FFT}^{-1}\{G(\alpha + j\omega_k)\}e^{\alpha k T_s}\}_{l=0} \tag{119}$$

where FFT $\{\cdot\}_{l=0}$ denotes the first element corresponding to $\omega_0 = 0$ in the resulting FFT series.

It is seen from (119) that the process static gain can be separately derived from the other model parameters, regardless of whether the relay test is unbiased or biased.

Note that there exists $G(\alpha + j\omega_1) = G(\alpha + j\omega_2)$ in the case where $\omega_2 > \omega_1$ while $P_u(\omega_2 - \omega_1) = 2h\pi$ and $\omega_2 - \omega_1 = 2l\pi/T_s$ are satisfied, and both h and l are positive integers. Frequency response estimation from (117) is therefore limited not only by the sampling frequency for computing the numerical integrals in (117), but also by the frequency range of $\omega \in [0, 2\pi \cdot 10^r)$, where r is the minimal integer that satisfies $10^r P_u = h$.

To cope with measurement noise, it is suggested to use 10–20 periods in the steady oscillation for the frequency response estimation. That is, the averaged frequency response estimation can be computed by

$$G(j\bar{\omega}_u) = \frac{\int_{t_{os}}^{t_{os}+N_s\bar{P}_u} y(t)e^{-j\bar{\omega}_u t}\,dt}{\int_{t_{os}}^{t_{os}+N_s\bar{P}_u} u(t)e^{-j\bar{\omega}_u t}\,dt} = \bar{A}_u e^{j\bar{\varphi}_u} \tag{120}$$

$$G(\alpha+j\omega) = \frac{(1-e^{-\bar{P}_u(\alpha+j\omega)})\int_0^{t_{os1}+N_1\bar{P}_u} [y(t)e^{-\alpha t}]e^{-j\omega t}\,dt + \int_{t_{os1}+N_1\bar{P}_u}^{t_{os1}+(N_1+1)\bar{P}_u} [y(t)e^{-\alpha t}]e^{-j\omega t}\,dt}{(1-e^{-\bar{P}_u(\alpha+j\omega)})\int_0^{t_{os1}+N_1\bar{P}_u} [u(t)e^{-\alpha t}]e^{-j\omega t}\,dt + \int_{t_{os1}+N_1\bar{P}_u}^{t_{os1}+(N_1+1)\bar{P}_u} [u(t)e^{-\alpha t}]e^{-j\omega t}\,dt} \tag{121}$$

where $\bar{\omega}_u = 2\pi/\bar{P}_u$, N_s is the number of steady oscillation periods used for averaging, and N_1 may be taken as large as possible but subject to a numerical constraint similar to (118) for computation.

Based on the estimated frequency response, the corresponding model fitting algorithms in the literature as aforementioned can be applied in terms of a specified model structure like FOPDT or SOPDT. For illustration, the state-of-the-art model fitting algorithm [59] in terms of the general transfer function model in (8) is briefly presented as below.

Considering that the low frequency range is primarily concerned for controller tuning in practice, one can establish multiple frequency response fitting conditions by choosing $\omega_M = (1.1\text{–}2.0)\omega_u$ and $\rho_k = \eta^k / \sum_{k=0}^{M} \eta^k$ to emphasize frequency response fitting over the low frequency range, where $M \in [5, 20]$ and $\eta \in [0.9, 0.99]$. To minimize the fitting error, a weighted LS objective function can be adopted as shown in (49).

To establish a linear regression for parameter estimation, assume that the process model is obtained as $\widehat{G}^{[i-1]}(s)$ at the $(i-1)$th iteration step, e.g. a model obtained from a prior knowledge of the process or a rough estimation based on a low-order model identification algorithm as presented in the previous sections, one can express the model to be derived at the ith iteration step using the multivariable Taylor series as

$$\begin{aligned}\widehat{G}^{[i]}(s) &= \widehat{G}^{[i-1]}(s) + \left.\frac{dG_n}{da_1}\right|_{\widehat{G}^{[i-1]}}(a_1^{[i]} - a_1^{[i-1]}) + \cdots + \left.\frac{dG_n}{da_n}\right|_{\widehat{G}^{[i-1]}}(a_n^{[i]} - a_n^{[i-1]}) \\ &+ \left.\frac{dG_n}{db_0}\right|_{\widehat{G}^{[i-1]}}(b_0^{[i]} - b_0^{[i-1]}) + \cdots + \left.\frac{dG_n}{db_m}\right|_{\widehat{G}^{[i-1]}}(b_m^{[i]} - b_m^{[i-1]}) + \left.\frac{dG_n}{d\theta}\right|_{\widehat{G}^{[i-1]}}(\theta^{[i]} - \theta^{[i-1]}) \\ &= \widehat{G}^{[i-1]}(s) + H^T(\gamma^{[i]} - \gamma^{[i-1]})\end{aligned} \tag{122}$$

where

$$\gamma^{[i]} = [a_1^{[i]}, \ldots, a_n^{[i]}, b_0^{[i]}, \ldots, b_m^{[i]}, \theta^{[i]}]^T \tag{123}$$

$$H = \left[\left.\frac{dG_n}{da_1}\right|_{\widehat{G}^{[i-1]}}, \ldots, \left.\frac{dG_n}{da_n}\right|_{\widehat{G}^{[i-1]}}, \left.\frac{dG_n}{db_0}\right|_{\widehat{G}^{[i-1]}}, \ldots, \left.\frac{dG_n}{db_m}\right|_{\widehat{G}^{[i-1]}}, \left.\frac{dG_n}{d\theta}\right|_{\widehat{G}^{[i-1]}} \right]^T \tag{124}$$

$$\left.\frac{dG_n}{da_k}\right|_{\widehat{G}^{[i-1]}} = \frac{b_m^{[i-1]}s^m + b_{m-1}^{[i-1]}s^{m-1} + \cdots b_1^{[i-1]}s + b_0^{[i-1]}}{(a_n^{[i-1]}s^n + a_{n-1}^{[i-1]}s^{n-1} + \cdots a_1^{[i-1]}s + 1)^2} s^k e^{-\theta^{[i-1]}s}, \quad k = 1, 2, \ldots, n \tag{125}$$

$$\left.\frac{dG_n}{db_l}\right|_{\widehat{G}^{[i-1]}} = \frac{s^l}{a_n^{[i-1]}s^n + a_{n-1}^{[i-1]}s^{n-1} + \cdots a_1^{[i-1]}s + 1} e^{-\theta^{[i-1]}s}, \quad l = 0, 1, 2, \ldots, m \tag{126}$$

$$\left.\frac{dG_n}{d\theta}\right|_{\widehat{G}^{[i-1]}} = -\frac{b_m^{[i-1]}s^{m+1} + b_{m-1}^{[i-1]}s^m + \cdots b_1^{[i-1]}s^2 + b_0^{[i-1]}s}{a_n^{[i-1]}s^n + a_{n-1}^{[i-1]}s^{n-1} + \cdots a_1^{[i-1]}s + 1} e^{-\theta^{[i-1]}s} \tag{127}$$

Let

$$Z(\alpha + j\omega_k) = G(\alpha + j\omega_k) + H^T(\alpha + j\omega_k)\gamma^{[i-1]} - \widehat{G}^{[i-1]}(\alpha + j\omega_k) \tag{128}$$

where $\omega_0 = 0$ and $k = 0, 1, 2, \ldots, M$.

Substituting (122) and (128) into the LS objective function in (49), one obtains an iterative objective function,

$$J_{\text{opt}}^{[i]} = \sum_{k=0}^{M} \rho_k \left| H^T(\alpha + j\omega_k)\gamma^{[i]} - Z(\alpha + j\omega_k) \right|^2 \tag{129}$$

Denote $\Phi = [H(\alpha + j\omega_1), \ldots, H(\alpha + j\omega_M)]^T$, $\Psi = [Z(\alpha + j\omega_1), \ldots, Z(\alpha + j\omega_M)]^T$, $W = diag\{\rho_0, \rho_1, \ldots, \rho_M, \rho_0, \rho_1, \ldots, \rho_M\}$, and

$$\bar{\Phi} = \begin{bmatrix} Re[\Phi] \\ Im[\Phi] \end{bmatrix}, \quad \bar{\Psi} = \begin{bmatrix} Re[\Psi] \\ Im[\Psi] \end{bmatrix}.$$

The iterative LS solution of (129) can be derived as

$$\gamma^{[i]} = (\bar{\Phi}^T W \bar{\Phi})^{-1} \bar{\Phi}^T W \bar{\Psi} \tag{130}$$

It can easily be verified that all the columns of $\bar{\Phi}$ are linearly independent from each other, so that $(\bar{\Phi}^T W \bar{\Phi})^{-1}$ is guaranteed to be non-singular for each iteration. Accordingly, the optimal solution can be derived in terms of a specified threshold of *ERR* in (49).

It should be noted that based on the Fourier expansion of the estimated process response around the oscillation frequency under a biased or unbiased relay test, model fitting conditions were also constructed in terms of the projection relationship between the Fourier series coefficients and the model parameters [89–91]. The achievable identification accuracy was not high due to the negligence of the higher order terms in the Fourier expansion, but in exchange for computation simplicity. Besides, the estimated process response under a relay test was used for estimating the frequency points of the closed-loop transfer function or sensitivity function for the convenience of closed-loop control design [92].

For illustration, the identification algorithms given in Refs. [56,58,59,90] are used to identify the first-, second-, and high-order processes (G_4–G_9) listed in Table 3, in comparison with the other categorized relay identification methods. Comparisons on the Nyquist fitting of G_8 and G_9 are shown in Figs. 10 and 11, respectively. It is seen that both the recent curve fitting approach [78,80] and the frequency response fitting method [58,59] can give good fitting accuracy. Compared to the identified FOPDT models, the identified SOPDT models yield obviously improved Nyquist fitting, especially for the high-order unstable process (G_9). To illustrate the achievable control performance for G_9, the SOPDT model identified by the frequency response fitting method [58] is adopted to apply the IMC-based 2DOF control method [8] for comparison with the describing function method [74] based on the same test as aforementioned. It is seen from the control results shown in Fig. 12 that obviously improved control performance is obtained by using the SOPDT model for control design. Note that the controller parameters ($\lambda_c = \lambda_f = 8$) cannot be tuned smaller for using the FOPDT model identified by the describing function method [74] in order to maintain the control system stability, but may be tuned much smaller to obtain faster output response based on the SOPDT model identified from either the curve fitting approach [80] or the frequency response fitting method [58].

4.4. Robust identification against measurement noise and load disturbance

Owing to the fact that measurement noise is commonly confronted in engineering practice, the use of an 'ideal' relay for an identification test is subject to relay chattering. To avoid measurement noise causing incorrect relay switching, it is generally suggested to use the relay hysteresis shown in (85). The magnitudes of ε_+ and ε_- should be set at least twice larger than the noise band, together with an upper limit almost equal to 0.95 times of the absolute minimum of u_+ or u_- [5]. After the process has moved into the working range, a short 'listening period' (e.g. 20–100 samples) should be referenced to set ε_+ and ε_- properly. Another strategy for maintaining correct relay switching was proposed in [93] by computing the time integral of the output response for reference, and furthermore, a low- or band-pass filter was suggested in tandem with the relay function to improve the denoising effect [94].

For using the noisy limit cycle data for model identification, if the noise level is low (e.g. NSR < 10%), the statistical averaging method can be used based on 5–20 steady oscillation periods to determine the limit cycle data for computation. To deal with a higher noise level, a low-pass Butterworth filter was suggested in [8] to recover the limit cycle under a relay test or carry out an offline denoising. The Butterworth filter can be determined by specifying the filter order, n_f, and the cutoff angular frequency, ω_c, that is,

$$\mathrm{Butter}(n_f, \omega_c) = \frac{b_1 + b_2 z^{-1} + b_3 z^{-2} + \cdots + b_{n_f+1} z^{-n_f}}{1 + a_2 z^{-1} + a_3 z^{-2} + \cdots + a_{n_f+1} z^{-n_f}} \tag{131}$$

where $\mathrm{Butter}(n_f, \omega_c)$ denotes the filtering function with two input parameters of n_f and ω_c.

Because measurement noise is mainly composed of high-frequency components in practical applications, a guideline for choosing the cutoff angular frequency was suggested [8] as

$$\omega_c \geq 5\omega_u = \frac{10\pi}{P_u} \tag{132}$$

Thereby, during a relay test only the measured output components below the frequency band regarding ω_u can be passed through for feedback control. Note that the phase lag caused by the low-pass filter almost does not affect measuring the oscillation period and the amplitude of the limit cycle, owing to that the relay output has the same phase lag under the filtered feedback control. Moreover, further improved denoising effect can be obtained by using an offline denoising strategy, that is, filtering the noisy limit cycle data in both the forward and reverse directions with the same low-pass Butterworth filter.

In the presence of load disturbance, by using an unbiased relay test, the influence can be intuitively detected by comparing the identity between the sequential periods of the limit cycle observed for computation. It was clarified in [95] that for the presence of a static load disturbance with a magnitude of h_d, there follows

$$\int_0^{P_u} h_d e^{-j\omega_u t}\, dt = h_d \int_0^{P_u} e^{-j\omega_u t}\, dt = 0 \tag{133}$$

Hence, computation of the process response at the oscillation frequency from (100) is not affected by a static load disturbance. Note that a static load disturbance may not be detected under a biased relay test, therefore causing identification errors.

By compensating for a static load disturbance in terms of using a relay test with asymmetrical relay output magnitudes, the recovered symmetrical limit cycle information was adopted for model identification [96]. Modified compensation algorithms were proposed in [97,98] for better estimating the ultimate gain and frequency for model identification. By decomposing a relay test into a sequence of step tests, Wang et al. [99] extended the time integral approach of step identification [26] to eliminate the influence from a static load disturbance and nonzero initial process conditions. In case there is a large static disturbance of which the magnitude is bigger than the relay magnitude, a PI controller was suggested [100] in tandem with the relay to reject the static disturbance, such that the recovered symmetrical output oscillation data could be used for model identification [101]. In contrast, by taking the magnitude of a static load disturbance as a parameter to be identified, the 'A-locus' method [78] was further extended in [81] to identify integrating processes. In addition, by computing the graphical areas of the limit cycle with respect to the time axis and using the Fourier series expansion of the relay response, the asymmetrical output response data were directly used for model identification under a static load disturbance [102].

4.5. Identification of multivariable processes

Owing to the merit of preventing the output response from drifting too far away from the setpoint, relay feedback tests have been explored for identifying MIMO processes. Generally, there are three choices of relay feedback implementation based on a multiloop control structure: (i) independent single-relay feedback (i.e. only one loop is under a relay test while the other loops are left open); (ii) sequential relay feedback (i.e. all loops will be closed one by one with respect to the sequence of relay tests while each relay test is for a single loop); (iii) decentralized relay feedback (i.e. all loops are simultaneously placed on relay feedback).

Due to the relay feedback and the corresponding interaction between individual loops, the closed-loop system stability must be taken into account for such an identification test. In fact, it is difficult to analyze the interaction between individual loops under multiloop relay tests. A few necessary conditions to forming the limit cycles under decentralized 'ideal' relay tests were discussed in Refs. [103,104]. By comparison, some necessary conditions to forming the limit cycles by using unbiased relay with hysteresis and/or dead zones were reported in Refs. [105,106]. It was concluded [104,106] that decentralized relay tests could guarantee steady oscillations for MIMO stable processes with diagonal dominance in the transfer function matrix, but might fail for MIMO processes with strong coupling between individual loops.

For the use of independent single-relay tests, the aforementioned relay identification methods for SISO processes can be conveniently extended for application. For using sequential relay tests or decentralized relay tests, few identification methods have been reported in the literature. Based on sequential relay tests, the DF method was extended for identifying the transfer functions of the desired input-output pairs [103], and an FFT technique was adopted to estimate the output frequency responses of an MIMO process by which model fitting could follow to derive the transfer function matrix [107]. Based on decentralized relay tests, a frequency response estimation algorithm was developed [108] for MIMO stable processes by assuming that a common oscillation frequency exists for all the loops, and an alternative algorithm [109] was presented for estimating the frequency response matrix in a low frequency range of primary concern for MIMO controller tuning.

5. Identification of specific nonlinear process models

Many chemical processes have dynamic response characteristics consisting of linear and nonlinear nature in tandem. For example, a typical extractive distillation column involves a nonlinear vapor–liquid equilibrium and linear mixing, while in contrast, a chemical pH process involves linear mixing in a vessel and a nonlinear static gain representing the pH titration curve. For the former, a static nonlinear element precedes the linear subsystem, which is generally named as 'Hammerstein' type, while for the latter, a linear subsystem precedes the static nonlinear element, which is named as 'Wiener' type. For more complicated systems, combinations of these linear and nonlinear blocks are adopted in practice to construct a model of Hammerstein–Wiener or Wiener–Hammerstein type in order to properly describe the system dynamic characteristics.

Because of the nonlinearity that may be described in different ways, parameter estimation approaches vary distinctly with the above model structures. Therefore, a suitable model structure should be determined before proceeding with the parameter estimation. For the choice of model structure, Luyben and Eskinat [110] suggested to perform consecutive two or more relay tests with different relay heights together with two different known dynamic modules inserted in the loop, such that a Hammerstein-type model should be adopted if the relays displaced vertically from the origin could lead to symmetrical output oscillation, and in contrast, a Wiener-type model should be adopted if the relays shifted horizontally from the origin could result in symmetrical pulses to the process input. The idea was further explored by Huang et al. [111] to a systematic identification procedure shown in Fig. 14, where the first-stage relay test is performed with $\varepsilon_+ = \varepsilon_- = 0$ and $\Delta\mu = 0$, and the second-stage relay test is performed by adjusting u_+ or u_- to compensate for the effect of the process static nonlinear gain. P_+ denotes the half period corresponding to the relay output u_+, and P_- the other half period corresponding to u_-, while A_+ denotes the positive output amplitude in the limit cycle and A_- the negative amplitude. It is seen that almost all of the above nonlinear model structures can be determined by such an identification procedure. For the use of an unbiased relay with hysteresis to perform an identification test, a criterion for model structure selection was suggested in [112], i.e. a Wiener-type model should be adopted if $\left|u_+(P_+ - P_-)\right| < \varepsilon_+$, and otherwise, a Hammerstein-type model should be adopted.

For identifying a Hammerstein-type model, based on using a modified step test consisting of multiple step changes that are similar to the pseudo-random binary signal (PRBS) in the frequency components but with a longer duration for each step change, Sung [113] presented a two-step identification method which could separately identify the static nonlinear function and the linear dynamic subsystem. Owing to that the minimization of the final output prediction error in terms of the whole input–output sample data was adopted to derive the static nonlinear function, asymptotic property on estimating either the static nonlinear function or the linear subsystem model could be quantitatively analyzed. For the use of a relay test, a so-called autotune variation (ATV) identification method was proposed by Luyben and Eskinat [110] where the DF method was used to identify the linear part. This method was applied in [114] for identifying a class of electrical drives, by which a controller autotuning method was successfully developed as demonstrated by experimental results. It was suggested by Park et al. [115] to implement the standard relay test followed by a triangular input test so as to estimate first the frequency response of the linear subsystem and then the nonlinear static function. Based on an LS fitting on the linear subsystem input sequence under these two tests, the identification of the nonlinear static function was separated from the linear subsystem model identification. By comparison, using the standard relay test followed by multiple pulse signals, Je et al. [116] proposed an improved plant response estimation to obtain more accurate data for identifying the linear subsystem, by pruning the relay oscillation harmonics as well as the influence from the input nonlinearity. To facilitate parameter estimation of a Hammerstein-type process, it was suggested [117] to insert an integrator in

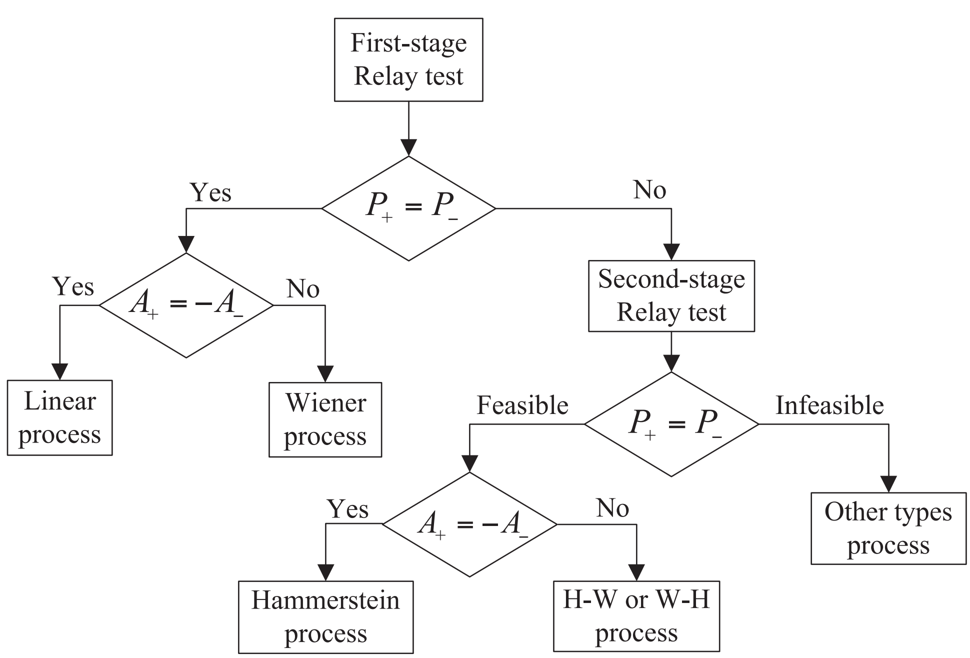

Fig. 14. Nonlinear model identification flow chart using relay tests.

the feedback loop to perform a relay test, such that the following relationship can be obtained for determining the static nonlinear function,

$$\left|\frac{v_+}{v_-}\right| = \frac{P_+}{P_-} \tag{134}$$

where $v_+ = N(u_+)$ and $v_- = N(u_-)$ denote the output values of the nonlinear block driven by the relay.

For identifying a Wiener-type model, using two step tests with different excitation magnitudes, Park and Lee [118] presented a two-step identification procedure: The first step is to identify the static nonlinear gain using all sampled data in the transient response, by introducing a small factor $(0 < \alpha < 1)$ to construct an LS fitting algorithm for estimating the intermediate output of the linear subsystem, and then deriving the linear model in terms of any developed step identification method as aforementioned. Based on a two-stage relay test, another two-step identification procedure was proposed by Huang et al. [119]: the first step is to identify the static nonlinear function by optimizing an objective function that deal with the asymmetry in the system output through adjusting the parameters of a predetermined nonlinear structure such as a rational polynomial. Upon determining the static nonlinear function, a sequence of the instrumental output is constructed in the second step for identifying the linear subsystem, i.e. the identification of a low-order model of FOPDT or SOPDT, for which many existing identification algorithms as aforementioned can be used. An extension of the above identification method was reported in [120] for improving parameter estimation, which can also be used to identify a Hammerstein-type model. In contrast, another two-step identification method was developed by Sung and Lee [121] based on using the Fourier expansion of the relay output to derive the static nonlinear gain for all samples in the limit cycle, and subsequently, a modified relay test to forming symmetric oscillation was suggested in [122] to eliminate the influence from a static load disturbance.

It should be noted that for MIMO nonlinear processes, step or relay identification methods for obtaining a Hammerstein- or Wiener-type model have been rarely reported in the literature, perhaps due to more difficulties and complexities involved with the model structure identification and parameter estimation, as discussed in the recent literature [123–127].

6. Concluding remarks and outlook

For industrial and chemical process identification using step or relay tests, the major technical approaches and identification methods developed in the past three decades have been surveyed in this paper. In the literature low-order models plus time delay such as FOPDT and SOPDT have been mostly adopted for identifying various industrial processes, owing to the convenience for control system design and controller tuning. The corresponding identification methods reported in the existing references as cited have been categorized in terms of the developed technical routes for overview. For using step tests, the main technical routes are classified into three groups, namely, the use of representative points in the transient response for model fitting, the time integral approach, and the use of frequency response estimation for model fitting. For using relay test(s), the main technical routes are also classified into three groups, namely, the describing function method, the curve fitting approach, and the use of frequency response estimation for model fitting. In addition, two specific categories are classified for robust step or relay identification methods coping with nonzero initial process conditions and/or load disturbance. The rationales of these categorized identification methods have been interpreted, respectively, through presenting a representative or state-of-the-art identification algorithm of each category. Simulation comparisons through benchmark examples on these representative algorithms have been made to illustrate the achievable identification accuracy, computation effort, and robustness against measurement noise and/or load disturbance, therefore facilitating the readers to have a general understanding of both the advantage and disadvantage of each categorized identification method as reviewed in chronological sequence.

For identifying nonlinear processes, the existing references have been mainly devoted to the identification of a Hammerstein- or Wiener-type model based on using multiple or modified step or relay tests. It has been recognized that the determination of a suitable model structure is of paramount importance for identifying such a nonlinear process. There is a common two-step identification procedure in the literature, one step for identifying the static nonlinear function, and the other step for identifying the linear dynamic subsystem for which the time delay model identification methods as reviewed can in general be adopted.

Though this survey has attempted to summarize the main contributions of step and relay identification methods developed in the past three decades for various industrial applications, there are still a lot of challenging issues regarding step or relay identification to be explored or solicited for advanced control system design and online autotuning. The following research topics are therefore pointed out to draw attention for future exploration:

(i) Robust frequency response estimation in the middle- or even high-frequency range required for advanced control of some industrial processes, e.g. the electric arc furnaces with heavy harmonic components, especially for the identification of high-order or very specific models with more parameters to describe complex system dynamics. Modified step or relay tests that are as simple as possible for implementation are desired for developing such frequency response estimation methods, especially for the presence of measurement noise.

(ii) Consistent parameter estimation for model identification using multiple step or relay tests subject to nonstatic or time-varying load disturbance, including a transient-type disturbance that exists only for short duration.

(iii) Asymptotic variance analysis for consistent parameter estimation based on a finite sampled data length or using multiple step or relay tests subject to measurement noise and/or load disturbance.

(iv) Model structure selection and optimal fitting criteria for piecewise linear model identification of nonlinear processes, and also for linearized model identification subject to deterministic (or inherent-type) load disturbance under different operating conditions.

(v) Online recursive step or relay identification algorithms for adaptive or gain-scheduling control design to realize real-time optimization, together with the model structure selection criteria and convergence analysis.

(vi) Closed-loop step test design for identifying SISO and MIMO integrating and unstable processes. Prior knowledge of the process should be minimized. Guidelines for guaranteeing the closed-loop stability together with the stability criteria are solicited for different applications.

(vii) Stability analysis for relay identification of various unstable processes, since little result has been cultivated in the literature as surveyed.
(viii) Design of relay feedback test for sustainable oscillations of an MIMO process that lead to the limit cycles, together with the stability conditions for relay identification of MIMO processes in terms of a multiloop or decoupling control structure.
(ix) Step or relay identification of non-square MIMO processes subject to no diagonal dominance, together with the criteria of choosing the input-output pairing for model identification.
(x) Criteria of model structure selection for identifying nonlinear processes of Hammerstein- or Wiener-type or the combination, from simple or modified step or relay tests, together with robust parameter estimation algorithms against measurement noise and/or load disturbance.

Acknowledgements

This work is supported in part by the Chinese Recruitment Program of 1000 Global Young Experts, the Fundamental Research Funds for the Central Universities of China, and the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program.

References

[1] D.P. Atherton, Oscillations in relay systems, Trans. Inst. Meas. Control (London) 3 (1982) 171–184.
[2] Y.Z. Tsypkin, Relay Control System, Cambridge University Press, Oxford, UK, 1984.
[3] K.J. Åström, T. Hägglund, Automatic tuning of simple regulators with specification on phase angle and amplitude margins, Automatica 20 (1984) 645–651.
[4] W.L. Luyben, Derivation of transfer functions for highly nonlinear distillation columns, Ind. Eng. Chem. Res. 26 (1987) 2490–2495.
[5] Q.G. Wang, T.H. Lee, C. Lin, Relay Feedback: Analysis, Identification and Control, Springer, London, 2003.
[6] C.C. Yu, Autotuning of PID Controllers: A Relay Feedback Approach, 2nd ed., Springer-Verlag, London, 2006.
[7] S.W. Sung, J. Lee, I.B. Lee, Process Identification and PID Control, Wiley, Singapore, 2009.
[8] T. Liu, F. Gao, Industrial Process Identification and Control Design: Step-test and Relay-experiment-based Methods, Springer, London, 2012.
[9] H. Rake, Step response and frequency response methods, Automatica 16 (1980) 519–526.
[10] K.M. Zhou, J.C. Doyle, K. Glover, Robust and Optimal Control, Prentice Hall, Englewood Cliff, NJ, 1996.
[11] D.E. Seborg, T.F. Edgar, D.A. Mellichamp, Process Dynamics and Control, 2nd ed., John Wiley & Sons, New Jersey, 2004.
[12] K.J. Åström, T. Hägglund, PID Controller: Theory, Design, and Tuning, 2nd ed., ISA Society of America, Research Triangle Park, NC, 1995.
[13] D.P. Atherton, Relay autotuning: an overview and alternative approach, Ind. Eng. Chem. Res. 45 (2006) 4075–4080.
[14] C.C. Hang, K.J. Åström, Q.G. Wang, Relay feedback auto-tuning of process controllers—a tutorial review, J. Process Control 12 (2002) 143–163.
[15] W.X. Zheng, Identification of closed-loop systems with low-order controllers, Automatica 32 (12) (1996) 1753–1757.
[16] T.E. Marlin, Process Control. Designing Processes and Control System for Dynamic Performance, McGraw-Hill, New York, 1995.
[17] J. Mikleš, M. Fikar, Process Modelling, Identification, and Control, Springer, Berlin, 2007.
[18] C.T. Huang, W.C. Clements, Parameter estimation for the second-order-plus-dead-time model, Ind. Eng. Chem. Process Des. Dev. 21 (1982) 601–603.
[19] G.P. Rangaiah, P.R. Krishnaswamy, Estimating second-order dead time parameters, Ind. Eng. Chem. Res. 33 (1994) 1867–1871.
[20] G.P. Rangaiah, P.R. Krishnaswamy, Estimating second-order dead time parameters from underdamped process transients, Chem. Eng. Sci. 51 (1996) 1149–1155.
[21] H.P. Huang, M.W. Lee, C.L. Chen, A system of procedures for identification of simple models using transient step response, Ind. Eng. Chem. Res. 40 (2001) 1903–1915.
[22] Q. Bi, W.J. Cai, E.L. Lee, Q.G. Wang, C.C. Hang, Y. Zhang, Robust identification of first-order plus dead-time model from step response, Control Eng. Pract. 7 (1999) 71–77.
[23] T. Söderström, P. Stoica, System Identification, Prentice Hall, New York, 1989.
[24] L. Ljung, System Identification: Theory for the User, 2nd ed., Prentice Hall, Englewood Cliff, NJ, 1999.
[25] Q.G. Wang, X. Guo, Y. Zhang, Direct identification of continuous time delay systems from step responses, J. Process Control 11 (2001) 531–542.
[26] Q.G. Wang, Y. Zhang, Robust identification of continuous systems with dead-time from step responses, Automatica 37 (2001) 377–390.
[27] T. Liu, K. Yao, F. Gao, Identification and autotuning of temperature control system with application to injection molding, IEEE Trans. Control Syst. Technol. 17 (2009) 1282–1294.
[28] M. Morari, E. Zafiriou, Robust Process Control, Prentice Hall, Englewood Cliff, NJ, 1989.
[29] F.G. Shinskey, Process Control System, 4th ed., McGraw Hill, New York, 1996.
[30] G.C. Goodwin, L. Graebe, M.E. Salgado, Control System Design, Prentice Hall, Upper Saddle River, NJ, 2001.
[31] K.J. Åström, T. Hägglund, Advanced PID Control, ISA Society of America, Research Triangle Park, NC, 2005.
[32] T. Liu, F. Gao, A frequency domain step response identification method for continuous-time processes with time delay, J. Process Control 20 (2010) 800–809.
[33] R. Pintelon, J. Schoukens, System Identification: A Frequency Domain Approach, IEEE Press, New York, 2001.
[34] S. Ahmed, B. Huang, S.L. Shah, Identification from step responses with transient initial conditions, J. Process Control 18 (2008) 121–130.
[35] S. Ahmed, B. Huang, S.L. Shah, Novel identification method from step response, Control Eng. Pract. 15 (2007) 545–556.
[36] Q.G. Wang, M. Liu, C.C. Hang, Y. Zhang, Y. Zhang, W.X. Zheng, Integral identification of continuous-time delay systems in the presence of unknown initial conditions and disturbances from step tests, Ind. Eng. Chem. Res. 47 (2008) 4929–4936.
[37] T. Liu, F. Gao, Robust step-like identification of low order process model under nonzero initial conditions and disturbance, IEEE Trans. Automat. Control 53 (2008) 2690–2695.
[38] S.H. Hwang, S.T. Lai, Use of two-stage least-squares algorithms for identification of continuous systems with time delay based on pulse responses, Automatica 40 (2004) 1561–1568.
[39] M. Liu, Q.G. Wang, B. Huang, C.C. Hang, Improved identification of continuous-time delay processes from piecewise step tests, J. Process Control 17 (2007) 51–57.
[40] T. Liu, F. Zhou, Y. Yang, F. Gao, Step response identification under inherent-type load disturbance with application to injection molding, Ind. Eng. Chem. Res. 49 (2010) 11572–11581.
[41] T. Liu, F. Gao, Closed-loop step response identification of integrating and unstable processes, Chem. Eng. Sci. 65 (2010) 2884–2895.
[42] S.Y. Li, W.J. Cai, H. Mei, Q. Xiong, Robust decentralized parameter identification for two-input two-output process from closed-loop step responses, Control Eng. Pract. 13 (2005) 519–531.
[43] M. Yuwana, D.E. Seborg, A new method for on-line controller tuning, AIChE J. 28 (1982) 434–440.
[44] I. Ananth, M. Chidambaram, Closed-loop identification of transfer function model for unstable systems, J. Franklin Inst. 336 (1999) 1055–1061.
[45] R.P. Sree, M. Chidambaram, Improved closed loop identification of transfer function model for unstable systems, J. Franklin Inst. 343 (2006) 152–160.
[46] C. Grimholt, S. Skogestad, Optimal PI-control and verification of the SIMC tuning rule, in: Proceedings IFAC Conference on Advances in PID Control, Brescia, Italy, March 28–30, 2012, ThPl.1 1–12.
[47] L. Wang, W.R. Cluett, System identification based on closed-loop step response data, IEE Proc. Control Theory Appl. 141 (1994) 107–110.
[48] H.P. Jin, I.P. Heung, I.-B. Lee, Closed-loop on-line process identification using a proportional controller, Chem. Eng. Sci. 53 (1998) 1713–1724.
[49] W.J. Cai, Y. Fang, Y.G. Wang, Robust closed-loop process identification from step test, in: The 8th International Conference on Control, Automation, Robotics and Vision, Kunming, China, December 6–9, 2004, pp. 907–912.
[50] P.N. Paraskevopoulos, G.D. Pasgianos, K.G. Arvanitis, New tuning and identification methods for unstable first order plus dead-time, IEEE Trans. Control Syst. Technol. 12 (2004) 455–464.
[51] E. Cheres, Parameter estimation of an unstable system with a PID controller in a closed loop configuration, J. Franklin Inst. 343 (2006) 204–209.
[52] T. Liu, Closed-loop step response identification for improving on-line autotuning of load disturbance rejection, in: The 30th Chinese Control Conference (CCC), Yantai, China, July 22–24, 2011, pp. 6328–6333.
[53] J.G. Ziegler, N.B. Nichols, Optimum settings for automatic controllers, Trans. ASME 64 (1942) 759–768.
[54] Q.G. Wang, C.C. Hang, B. Zou, Low-order modeling from relay feedback, Ind. Eng. Chem. Res. 36 (1997) 375–381.

[55] R.C. Panda, C.C. Yu, Analytical expressions for relay feed back responses, J. Process Control 13 (2003) 489–501.
[56] T. Liu, F. Gao, Alternative identification algorithms for obtaining a first-order stable/unstable process model from a single relay feedback test, Ind. Eng. Chem. Res. 47 (2008) 1140–1149.
[57] T. Liu, F. Gao, Y.Q. Wang, A systematic approach for on-line identification of second-order process model from relay feedback test, AIChE J. 54 (2008) 1560–1578.
[58] T. Liu, F. Gao, Identification of integrating and unstable processes from relay feedback, Comput. Chem. Eng. 32 (2008) 3038–3056.
[59] T. Liu, F. Gao, A generalized approach for relay identification of time delay and nonminimum phase processes, Automatica 45 (2009) 1072–1079.
[60] K.K. Tan, Q.G. Wang, T.H. Lee, Finite spectrum assignment control of unstable time delay processes with relay tuning, Ind. Eng. Chem. Res. 37 (1998) 1351–1357.
[61] S.H. Shen, H.D. Yu, C.C. Yu, Autotune identification for systems with right-half-plane poles and zeros, J. Process Control 9 (1999) 161–169.
[62] M. Majhi, D.P. Atherton, Online tuning of controllers for an unstable FOPDT process, IEE Proc. Control Theory Appl. 147 (2000) 421–427.
[63] T. Thyagarajan, C.C. Yu, Improved autotuning using the shape factor from relay feedback, Ind. Eng. Chem. Res. 42 (2003) 4425–4440.
[64] C. Lin, Q.G. Wang, T.H. Lee, Relay feedback: a complete analysis for first-order systems, Ind. Eng. Chem. Res. 43 (2004) 8400–8402.
[65] T. Co, Relay stabilization and bifurcations of unstable SISO processes with time delay, IEEE Trans. Automat. Control 55 (2010) 1131–1141.
[66] P.K. Padhy, S. Majhi, Relay based PI–PD design for stable and unstable FOPDT processes, Comput. Chem. Eng. 30 (2006) 790–796.
[67] W.L. Luyben, Getting more information from relay feedback tests, Ind. Eng. Chem. Res. 40 (2001) 4391–4402.
[68] R.C. Panda, C.C. Yu, Shape factor of relay response curves and its use in autotuning, J. Process Control 15 (2005) 893–906.
[69] W. Li, E. Eskinat, W.L. Luyben, An improved autotune identification method, Ind. Eng. Chem. Res. 30 (1991) 1530–1541.
[70] H.P. Huang, J.C. Jeng, K.Y. Luo, Auto-tune system using single-run relay feedback test and model-based controller design, J. Process Control 15 (2005) 713–727.
[71] A. Leva, PID autotuning algorithm based on relay feedback, IEE Proc. Control Theory Appl. 140 (1993) 328–338.
[72] S.H. Shen, J.S. Wu, C.C. Yu, Use of biased-relay feedback for system identification, AIChE J. 42 (1996) 1174–1180.
[73] G. Marchetti, C. Scali, Use of modified relay techniques for the design of model-based controllers for chemical processes, Ind. Eng. Chem. Res. 39 (2000) 3325–3334.
[74] G. Marchetti, C. Scali, D.R. Lewin, Identification and control of open-loop unstable processes by relay methods, Automatica 37 (2001) 2049–2055.
[75] P. Wang, D.Y. Gu, W.D. Zhang, Modified relay feedback identification based on describing function analysis, Ind. Eng. Chem. Res. 46 (2007) 1538–1546.
[76] H.P. Huang, C.L. Chen, C.W. Lai, G.B. Wang, Autotuning for model-based PID controllers, AIChE J. 42 (1996) 2687–2691.
[77] D.Y. Gu, L.L. Ou, P. Wang, W.D. Zhang, Relay feedback autotuning method for integrating processes with inverse response and time delay, Ind. Eng. Chem. Res. 45 (2006) 3119–3132.
[78] I. Kaya, D.P. Atherton, Parameter estimation from relay autotuning with asymmetric limit cycle data, J. Process Control 11 (2001) 429–439.
[79] S. Majhi, Relay-based identification of a class of non-minimum phase SISO processes, IEEE Trans. Automat. Control 52 (2007) 134–139.
[80] S. Majhi, Relay based identification of processes with time delay, J. Process Control 17 (2007) 93–101.
[81] I. Kaya, Parameter estimation for integrating processes using relay feedback control under static load disturbances, Ind. Eng. Chem. Res. 45 (2006) 4726–4731.
[82] Q. Bi, Q.G. Wang, C.C. Hang, Relay-based estimation of multiple points on process frequency response, Automatica 33 (1997) 1753–1757.
[83] M. Friman, K.V. Waller, A two-channel relay for autotuning, Ind. Eng. Chem. Res. 36 (1997) 2662–2671.
[84] C.H. Jeon, Y.J. Cheon, J.S. Kim, S.W. Sung, J. Lee, Relay feedback methods combining sub-relays to reduce harmonics, J. Process Control 20 (2010) 228–234.
[85] Q.G. Wang, C.C. Hang, Q. Bi, A technique for frequency response identification from relay feedback, IEEE Trans. Control Syst. Technol. 7 (1999) 122–128.
[86] S.W. Sung, I.B. Lee, An improved algorithm for automatic tuning of PID controllers, Chem. Eng. Sci. 55 (2000) 1883–1891.
[87] M. Ma, X. Zhu, A simple auto-tuner in frequency domain, Comput. Chem. Eng. 30 (2006) 581–586.
[88] Q.G. Wang, C.C. Hang, Q. Bi, Process frequency response estimation from relay feedback, Control Eng. Pract. 5 (1997) 1293–1302.
[89] K. Srinivasan, M. Chidambaram, Modified relay feedback method for improved system identification, Comput. Chem. Eng. 27 (2003) 727–732.
[90] V. Ramakrishnan, M. Chidambaram, Estimation of a SOPTD transfer function model using a single asymmetrical relay feedback test, Comput. Chem. Eng. 27 (2003) 1779–1784.
[91] S. Vivek, M. Chidambaram, Identification using single symmetrical relay feedback test, Comput. Chem. Eng. 29 (2005) 1625–1630.
[92] G.H.M. de Arruda, P.R. Barros, Relay-based closed-loop transfer function frequency points estimation, Automatica 39 (2003) 309–315.
[93] J.S. Kim, J. Byeon, D. Chun, S.W. Sung, J. Lee, Relay feedback method for processes under noisy environments, AIChE J. 56 (2010) 560–562.
[94] J. Lee, J.S. Kim, J. Byeon, S.W. Sung, T.F. Edgar, Relay feedback identification for processes under drift and noisy environments, AIChE J. 57 (2011) 1809–1816.
[95] Q.G. Wang, H.W. Fung, Y. Zhang, Robust estimation of process frequency response from relay feedback, ISA Trans. 38 (1999) 3–9.
[96] C.C. Hang, K.J. Åström, W.K. Ho, Relay auto-tuning in the presence of static load disturbance, Automatica 29 (1993) 563–564.
[97] S.H. Shen, S.J. Wu, C.C. Yu, Autotune identification under load disturbance, Ind. Eng. Chem. Res. 35 (1996) 1642–1651.
[98] J.H. Park, S.W. Sung, I.B. Lee, Improved relay auto-tuning with static load disturbance, Automatica 33 (1997) 711–715.
[99] Q.G. Wang, M. Liu, C.C. Hang, W. Tang, Robust process identification from relay tests in the presence of nonzero initial conditions and disturbance, Ind. Eng. Chem. Res. 45 (2006) 4063–4070.
[100] S.W. Sung, J. Lee, Relay feedback method under large static disturbances, Automatica 42 (2006) 353–356.
[101] S.W. Sung, J. Lee, D.H. Lee, J.H. Han, Y.S. Park, Two-channel relay feedback methods under static disturbances, Ind. Eng. Chem. Res. 45 (2006) 4071–4074.
[102] J. Lee, S.W. Sung, T.F. Edgar, Area method for a biased relay feedback system, Ind. Eng. Chem. Res. 49 (2010) 8016–8020.
[103] A.P. Loh, C.C. Hang, C.K. Quek, V.U. Vasnani, Autotuning of multiloop proportional-integral controllers using relay feedback, Ind. Eng. Chem. Res. 32 (1993) 1102–1107.
[104] A.P. Loh, V.U. Vasnani, Necessary conditions for limit cycles in multiloop relay systems, IEE Proc. Control Theory Appl. 141 (1994) 163–168.
[105] Z.J. Palmor, Y. Halevi, N. Krasney, Limit cycles in decentralized relay systems, Int. J. Control 56 (1992) 755–765.
[106] Z.J. Palmor, Y. Halevi, N. Krasney, A general and exact method for determining limit cycles in decentralized relay systems, Automatica 31 (1995) 1333–1339.
[107] Q.G. Wang, Y. Zhang, A novel FFT-based robust multivariable process identification method, Ind. Eng. Chem. Res. 40 (2001) 2485–2494.
[108] Q.G. Wang, B. Zou, T.H. Lee, Q. Bi, Auto-tuning of multivariable PID controllers from decentralized relay feedback, Automatica 33 (1997) 319–330.
[109] Y.G. Wang, W.J. Cai, M. Ge, Decentralized relay-based multivariable process identification in the frequency domain, IEEE Trans. Automat. Control 48 (2003) 873–877.
[110] W. Luyben, E. Eskinat, Nonlinear auto-tune identification, Int. J. Control 59 (1994) 595–626.
[111] H.P. Huang, M.W. Lee, C.Y. Tsai, Structure identification of block-oriented nonlinear models using relay feedback tests, J. Chem. Eng. Jpn. 34 (2001) 748–756.
[112] U. Mehta, S. Majhi, Identification of a class of Wiener and Hammerstein-type nonlinear processes with monotonic static gains, ISA Trans. 49 (2010) 501–509.
[113] S.W. Sung, System identification method for Hammerstein processes, Ind. Eng. Chem. Res. 41 (2002) 4295–4302.
[114] A. Balestrino, A. Landi, M. Ould-Zmirli, L. Sani, Automatic nonlinear auto-tuning method for Hammerstein modeling of electrical drives, IEEE Trans. Ind. Electron. 48 (2001) 645–655.
[115] H.C. Park, D.G. Koo, J.H. Youn, J. Lee, S.W. Sung, Relay feedback approaches for the identification of Hammerstein-type nonlinear processes, Ind. Eng. Chem. Res. 43 (2004) 735–740.
[116] C.H. Je, J. Lee, S.W. Sung, D.H. Lee, Enhanced process activation method to remove harmonics and input nonlinearity, J. Process Control 19 (2009) 353–357.
[117] J.C. Jeng, M.W. Lee, H.P. Huang, Identification of block-oriented nonlinear processes using designed relay feedback tests, Ind. Eng. Chem. Res. 44 (2005) 2145–2155.
[118] H.C. Park, J. Lee, Step and pulse response methods for identification of Wiener processes, AIChE J. 52 (2006) 668–677.
[119] H.P. Huang, M.W. Lee, Y.T. Tang, Identification of Wiener model using relay feedback test, J. Chem. Eng. Jpn. 31 (1998) 604–612.
[120] M.W. Lee, H.P. Huang, J.C. Jeng, Identification and controller design for nonlinear processes using relay feedback, J. Chem. Eng. Jpn. 37 (2004) 1194–1206.
[121] S.W. Sung, J. Lee, Modeling and control of Wiener-type processes, Chem. Eng. Sci. 59 (2004) 1515–1521.
[122] S.W. Sung, J. Lee, Relay feedback method under nonlinearity and static disturbance conditions, Ind. Eng. Chem. Res. 45 (2006) 4028–4031.
[123] Y. Zhu, Multivariable System Identification for Process Control, Elsevier Science, London, 2001.
[124] J.C. Jeng, H.P. Huang, Nonparametric identification for control of MIMO Hammerstein systems, Ind. Eng. Chem. Res. 47 (2008) 6640–6647.
[125] S.I. Biagiola, J.L. Figueroa, Identification of uncertain MIMO Wiener and Hammerstein models, Comput. Chem. Eng. 35 (2011) 2867–2875.
[126] X.M. Chen, H.F. Chen, Recursive identification for MIMO Hammerstein systems, IEEE Trans. Automat. Control 56 (2011) 895–902.
[127] C. Qi, H.X. Li, X. Zhao, S. Li, F. Gao, Hammerstein modeling with structure identification for multi-input multi-output nonlinear industrial processes, Ind. Eng. Chem. Res. 50 (2011) 11153–11169.