Benchmark problems for continuous-time model identification: Design aspects, results and perspectives

Valentin Pascu, Hugues Garnier, Lennart Ljung, Alexandre Janot

HAL Id: hal-02262375
https://hal.science/hal-02262375

Submitted on 2 Aug 2019

Benchmark Problems for Continuous-Time Model Identification: Design Aspects, Results and Perspectives [★]

Valentin Pascu [a,b], Hugues Garnier [b,c], Lennart Ljung [d], Alexandre Janot [a]

[a]*Systems Control and Flight Dynamics Department, ONERA - The French Aerospace Laboratory, 31055 Toulouse, France*

[b]*Université de Lorraine, CRAN, UMR 7039, 54519 Vandoeuvre-lès-Nancy, France*

[c]*CNRS, CRAN, UMR 7039, 54506 Vandoeuvre-lès-Nancy, France*

[d]*Division of Automatic Control, Linköping University, SE-58183 Linköping, Sweden*

Abstract

The problem of estimating continuous-time model parameters of linear dynamical systems using sampled time-domain input and output data has received considerable attention over the past decades and has been approached by various methods. The research topic also bears practical importance due to both its close relation to first principles modeling and equally to linear model-based control design techniques, most of them carried in continuous time. Nonetheless, as the performance of the existing algorithms for continuous-time model identification has seldom been assessed and, as thus far, it has not been considered in a comprehensive study, this practical potential of existing methods remains highly questionable. The goal of this brief paper is to bring forward a first study on this issue and to factually highlight the main aspects of interest. As such, an analysis is performed on a benchmark designed to be consistent both from a system identification viewpoint and from a control-theoretic one. It is concluded that robust initialization aspects require further research focus towards reliable algorithm development.

Key words: identification algorithms, output error identification, parameter identification, linear multivariable systems, benchmark examples, Monte Carlo simulation.

1 Introduction

The estimation of continuous-time models of dynamical systems from sampled input and output system data is a research topic that has known considerable advancements throughout the past two decades, currently having reached maturity. With relevance pertaining to the traditional modeling framework for physical systems using first principles, the need for continuous-time system identification methods has been acknowledged also from model-based control viewpoint, given that *prior* knowledge can be preserved and further used for control design. Consequently, the major overhaul of the MATLAB[®] System Identification Toolbox in 2012 [15] aimed at a rapprochement with the Control System and with the Robust Control toolboxes. Currently, the default use of many of the identification algorithms gives a continuous-time model while discrete-time models still can be optionally obtained. This reflects the fact that

for most users, continuous-time models remain more intuitive and estimating such models is a more natural step for the control community. Indeed, many advantages regarding the quality and versatility of the identified continuous-time models can also be achieved using direct continuous-time approaches [6]. While many developments within this area have been reported, an exhaustive assessment of the associated algorithms on a relevant test bench has not been documented.

Recently, within the system identification and the machine learning communities, growing attention is given to proposing suitable benchmarks for numerically assessing methodologies [10,11,23]. The value of a test bench depends on the community's acceptance of its relevance with respect to the different techniques available, as well as the data generating procedure and the model quality assessment criteria. Benchmarks based on randomly-generated systems, though currently common, remain under question with respect to their appropriateness and representativity for the issues at hand [21]. Nonetheless, the necessity for having such benchmarks based on ran-

domized systems available for various studies does exist from both a theoretical and practical point of view. In control engineering, most systems of practical interest obey the laws of physics and, therefore, have an inherent low-pass behaviour, even when accompanied by additional resonant/anti-resonant dynamics. Well-designed test benches based on randomized systems experiencing such behaviors can, therefore, be considered attractive from both the automatic control and the system identification viewpoint, while thorough studies based on them with respect to given questions at hand can but reveal important aspects for further development.

Though crucial in understanding the relevance of fundamental works with respect to envisioned applications, as well as to realizing their limitations, it often remains cumbersome to point out where methodologies fall short of their expected standard [19]. Yet, the challenge is approachable within the control and identification communities by use of computational methods for assessment [2,1], as is the case within this study. The goal of this brief paper is to qualify and quantify, from both these perspectives, the reliability of continuous-time Output-Error (OE) model identification algorithms on a custom-developed benchmark and to pertinently highlight the important aspects for future advances. The study leaps forward from the presentation given in [19] by bringing about a systematically designed test bench and drawing consistent conclusions using it in relation to these goals. The choice of model class is here restricted to OE models, these readily portraying the essential aspects of the work; more challenging and realistic test systems can also be conceived *e.g.* using other linear model classes (such as Box-Jenkins) or even nonlinear ones, yet this falls outside the scope of the current study.

The paper is organized as follows. In Section II, the identification of multi-input continuous-time transfer function models is briefly reviewed. In Section III, a control- and identification-relevant setup is developed for assessing the algorithm implementations of the standard approaches described in Section II. In Section IV, results obtained with the proposed benchmark are summarized and discussed. Based on these, specific conclusions are drawn in Section V and several perspectives are given.

2 Continuous-Time Transfer Function Model Identification in a Multi-Input Setting

Let us consider the general problem of estimating multiple-input single-output (MISO) continuous-time (CT) rational transfer function (TF) matrices, expressed in terms of the Laplace transform variable s, as:

$$G(s) = \left[\frac{B_1(s)}{A_1(s)} \, \frac{B_2(s)}{A_2(s)} \, \cdots \, \frac{B_l(s)}{A_l(s)} \right] \tag{1}$$

starting from N samples of the l-dimensional CT input

$u(t) = [u_1(t) \, u_2(t) \, ... \, u_l(t)]^T$ and output $y(t)$ signals, collected into the $Z^N = \{u(t_k), y(t_k)\}_{k=1}^N$ dataset. By taking the number of inputs to be $l = 1$, a single-input single-output (SISO) scalar rational transfer function is obtained. The input and output signals are related by:

$$y(t) = G(p)u(t) + v(t) \tag{2}$$

with $v(t)$ representing a CT signal incorporating the effects of measurement noise and output disturbances, whose effect on $y(t)$ is considered only at the sampling instants t_k, $k = 1, ..., N$ in the dataset Z^N. Let p denote the differential operator. The noise-free response of the system is defined as $\mathring{y}(t) \triangleq G(p)u(t)$. In this section we briefly recall two main approaches for solving this estimation problem, namely the *Maximum Likelihood* and the *Instrumental Variable* methods.

2.1 The Maximum Likelihood Method

With appropriate account taken of the intersample behaviour of $u(t)$ and $y(t)$ and under the assumption that $v(t)$ is a stochastic signal with Gaussian probability density function, the theoretically optimal solution is to apply the *Maximum Likelihood* method. It has long been known how to do this *e.g.* [17,12] and a recent discussion is given in [16]. If the disturbances on the system are Gaussian, the ML method coincides with the *Prediction Error Method*. For each input $i = 1, ..., l$ the parameters of $B_i(s)$ and $A_i(s)$ from (1) are collected in the vector $\theta = \left[\theta_1^T \, \theta_2^T \, \cdots \, \theta_l^T \right]$, so that the predicted θ-parametrized output is given as:

$$\hat{y}(t|\theta) = \sum_{i=1}^l G_i(p, \theta_i) u_i(t) \tag{3}$$

We denote the true parameter vector of $G(p)$ by θ_o. If in (2) the additive disturbance $v(t)$ at the output is white then the optimal estimate arises as the solution of:

$$\hat{\theta}_{ML} = \arg \min_\theta \sum_{k=1}^N \|y(t_k) - \hat{y}(t_k|\theta)\|^2 \tag{4}$$

The `tfest` algorithm from the MATLAB$^\circledR$ System Identification Toolbox [13] is a software implementation of this approach, further denoted by TFEST in the paper.

2.2 The Instrumental Variable Method

Another route to solving the formulated problem comes from the application of the *Instrumental Variable* method, see *e.g.* [26,27]. The parameter estimate then arises as the solution of:

$$\hat{\theta}_{IV} = \arg \min_\theta \sum_{k=1}^N \left\| \zeta_f(t_k) \left(y_f(t_k) - \varphi_f^T(t_k)\theta \right) \right\|^2 \tag{5}$$

2

where $\zeta_f(t_k)$ and $\varphi_f(t_k)$ are the filtered instrument and regression vectors, respectively and $y_f(t_k)$ is the filtered output signal [5,28]. The `srivc` algorithm of the CONTSID Toolbox [4] contains an implementation of this approach, further called SRIVC in the study.

3 Description of the Proposed Benchmark

An evaluation setup developed for investigating the performance of the algorithms that implement the presented methods, is described in this section. The key factors that lead to correctly-drawn conclusions on obtained results are the *generation of effective low-pass systems*, the *design of informative experiments* and the *choice of pertinent assessment criteria* and will here be considered sequentially.

3.1 Generating appropriate test systems

Given the assumptions resilient to the presented identification methods, a simulation setup that satisfies these as closely as possible is defined. First, a CT OE structure is chosen for (2) by setting the disturbance $v(t)$ to be a zero-mean white Gaussian signal. Reference systems of SISO and MISO type are taken in (1) with $l = \{1, 2\}$, where:

$$G_i(s) = \frac{\alpha_i \prod_{j=1}^m (s - z_j)}{\prod_{j=1}^n (s - p_j)}, \, i = 1, ..., l \qquad (6)$$

In general, the most appropriate way to avoid any specificity of the dynamics in the reference systems, apart from the low-pass behaviour, is to choose the transfer function parameters in a random way. However, typical related pitfalls, such as approximate pole-zero cancellation and generation of low-effective order systems, common to both single- and multiple-input system realization problems must be avoided [21].

For designing randomized low-pass systems of various orders n, strictly proper stable rational transfer functions with pole-zero excess of 1 are taken for $G_i(s)$ with their n poles and $m = n - 1$ zeros being randomly-chosen and *e.g.* evenly distributed in a given frequency band $\omega \in \left[10^{-3}\omega_B, \omega_B\right]$, where the lower limit is chosen such that the system dynamics are not too wide apart. For each system order, ω_B is split randomly into several divisions, as follows. One by one, a pole pair is designated to either consist of one real or two complex-conjugate poles; this results into a number $\eta \leq n$ of pole pairs, representing also the number of frequency divisions. The real part (for real poles) or the imaginary parts (for complex poles) are chosen such that the given pair has its associated frequency within the designated division, but chosen randomly based on *e.g.* a normal distribution. For complex poles, the real part is

also randomly-chosen on *e.g.* a normal distribution, in the interval $\left[-\omega_B, 10^{-3}\omega_B\right]$, without loss of generality.

Similarly, the zeros can be assigned in their own corresponding divisions $\mu \leq m$; the zero pair location with respect to the imaginary axis (left-half or right-half plane) is randomly-chosen, so as to allow for both minimum and non-minimum phase systems. The gain α_i is randomly-picked based on *e.g.* an integer uniform distribution so that the DC gains of the systems are above 0 dB and its sign is randomly-chosen; there is, again, no loss of generality in the choice. Generating random test systems of different complexities, expressed in terms of the order n, is done numerically. As a first guarantee for *effective model complexity*, cases where the locations of poles are close to those of zeros should be avoided by re-iteration when the distance is *e.g.* less than $\omega_B/(m + n)$. This goal can be achieved by various criteria, whereas the proposed one provides a simple method relating the frequency band of interest to the number of singularities. For simplicity, we further consider $\omega_B = 2\pi 100$ rad/s in the remainder of the study.

Indeed, the reference systems play an important role with respect to the thoroughness of the identification algorithm assessment. For a second guarantee on the effectiveness of the model complexity, the generated systems should be checked not to be of low-effective orders [21]. This can be evaluated from the computation of the Hankel singular values [9,29]. For example, the computation of the ratio $\overline{\sigma}/\underline{\sigma}$ between the largest $\overline{\sigma}$ and the lowest $\underline{\sigma}$ Hankel singular values represents up to some extent the degree of model reducibility. This allows systems of low-effective orders to be defined here as those systems for which $\overline{\sigma}/\underline{\sigma} < e^{l \cdot n}$, the right-hand side stemming from the fact that the decay rate in the Hankel singular values is exponential for asymptotically stable systems [18], as in the case considered here. If the condition is not satisfied, the system generation is re-iterated.

The summary of one set of generated SISO and MISO systems, designated as reference for the benchmark, is provided in Table 1. As an example of the proposed random system generation technique, the pole-zero map for the 5th order SISO system is given in Figure 1. As can be noticed, first a complex pole pair is generated, followed by a real pole and, subsequently, another complex pole pair, so here $\eta = 3$ divisions exist for the given frequency range; for the zeros one complex zero pair followed by two real ones are generated, corresponding to $\mu = 3$.

In Figure 2, the Bode diagrams for 25 SISO systems of order 5 are given: as can be seen from their frequency response functions, the procedure delivers stable low-pass SISO systems with diverse dynamics and DC gains.

When used for generating MISO systems, the proposed procedure remains efficient in delivering diverse multi-input models in a satisfactory manner, given the system

3

System	SISO		MISO	
Order	$l \cdot (m+n)$	$\overline{\sigma}/\underline{\sigma}$	$l \cdot (m+n)$	$\overline{\sigma}/\underline{\sigma}$
1	2	1.0e0	4	6.5e0
2	4	1.5e0	8	1.7e1
3	6	1.7e0	12	7.8e1
4	8	1.4e0	16	7.2e1
5	10	4.2e0	20	6.4e1
6	12	4.2e0	24	6.5e2
7	14	5.0e0	28	1.1e5
8	16	4.3e0	32	2.1e3
9	18	7.2e0	36	3.6e5
10	20	8.7e0	40	3.2e8

Table 1
Number of parameters and ratios of maximum/minimum
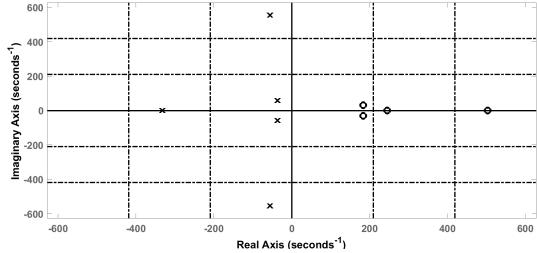Hankel singular values for the reference systems

Fig. 1. Pole-zero map for a 5th order SISO system; the three corresponding frequency divisions are highlighted
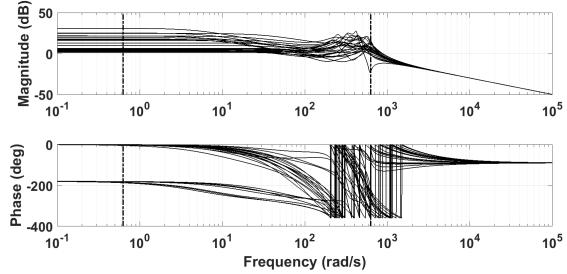
Fig. 2. Bode diagram for 25 SISO systems of 5th order; range $\left[10^{-3}\omega_B, \omega_B\right]$ for $\omega_B = 2\pi 100$ rad/s highlighted

dynamics shown on singular value plots of Figure 3 for the same number of 2x1 MISO models of 5th order.

3.2 Input design for informative experiments

These reference systems are to be simulated for collecting identification data. The input signals for excitation are chosen to be of random-phase multisine type:

$$u_i(t) = \sum_{\lambda=1}^{F} A_i \sin(2\pi f_\lambda t + \phi_\lambda), \; i = 1, 2 \qquad (7)$$
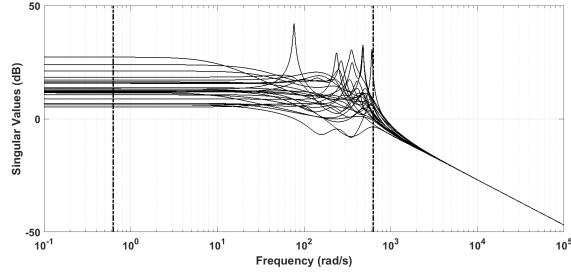
Fig. 3. Singular value plot for 25 2x1 MISO systems of 5th order; range $\left[10^{-3}\omega_B, \omega_B\right]$ for $\omega_B = 2\pi 100$ rad/s highlighted

as this class experiences the highest versatility with respect to a given designable experiment [20]. The design parameters of the multisine(s) need to be chosen such that the excitation is persistent for parametric estimation in a certain frequency range e.g. $\omega_R = 1.1\omega_B$. The choice of the number of harmonics F can either be made to be as large as possible or, at the least, not smaller than $\Omega = 2 \cdot (m+n)$ in relation to the number of parameters to be estimated for each system [12]. A choice of as high as possible a number of harmonics e.g. 1000 within the frequency range of interest has here been made, without any loss of generality. The harmonics ω_λ are linearly spaced from $\omega_0 = 10^{-3}\omega_B$ to ω_R, while for each harmonic λ, the phase ϕ_λ is chosen to be a uniformly-distributed random variable with zero mean and variance $\pi/2$. For ensuring appropriate informativity in MISO experiments input orthogonality is recommended, see e.g. [3]. In the 2x1 MISO case of the current study, this implies that for any number of periods, the second input will be a $\pi/2$ phase-shifted version of the first. An equally valid approach would be to use orthogonal multisines as defined in [3], but this is not opted for within the benchmark. Each multisine has a period of length $T_0 = 2\pi/\omega_0$. The spectrum of each multisine signal i can be chosen to be flat, by defining the same amplitude A_i for each harmonic; however, to ensure that the data is informative for multivariable system estimation [7], these amplitudes need to be chosen such that the inputs contributions to the output are comparable, e.g. by defining $A_1 = |G_2(0)/G_1(0)|$ and $A_2 = |G_1(0)/G_2(0)|$ for the MISO case and $A = 1$ for SISO; this often remains reasonable in practice given that the DC gain of systems under investigation are roughly known prior to experimentation. The inputs are fast-sampled, as is typically the case in CT identification, here with a frequency of $F_s = 50\frac{\omega_B}{2\pi}$ Hz.

The reference systems are simulated using e.g. `lsim` with a first-order hold specification and zero initial conditions. The number of periods for excitation is chosen such that sufficient data is available for estimation e.g. several periods. When choosing the number of periods for the experiment, attention must be given also to whether or not the identification method can handle transient data; if this is not the case, a certain amount of samples corre-

sponding to the transient's response must be discarded from the obtained data, while still allowing for sufficient data for estimation. Since the two methods presented in Section II are reasonably robust with respect to cases when transient data forms part of the provided identification data, any choice of data discarding can be conceived within the presented benchmark.

On the output, zero-mean white Gaussian measurement noise with average power P_v is added such that a signal-to-noise ratio SNR= $10 \log (P_{\mathring{y}}/P_v)$ of 10 dB is obtained in the data. Here $P_{\mathring{y}}$ is the average power of the noise-free output.

3.3 Criteria for systematic model quality evaluation

Model quality assessment can be carried in many different ways depending on the goal of the model, such as *prediction* [12] or *feedback control design* [25,24]. The model's prediction performance is often assessed by evaluation of standard statistics (such as mean and standard deviation or median and the associated first and third quartiles) of the time-domain fits between the model noise-free output $\mathring{y}(t)$ and the measured output $y(t)$ given by:

$$FIT_\gamma = 100 \cdot \left[1 - \frac{\|y(t_k) - \mathring{y}(t_k)\|_2}{\|y(t_k) - \mathrm{E}y(t_k)\|_2} \right], \gamma = 1, ..., \Gamma, \ (8)$$

offers for a Monte Carlo-type of analysis reliable statistics with respect to Γ, provided that Φ *estimation failures* (any case when the estimation algorithms fail to provide a parameter estimate) from the Γ obtained models have been counted, but not indexed as outliers in the calculation of the statistics.

Additionally, in any case where a parameter estimate has been obtained (an estimation failure did not occur), the parameter can be *a priori* expected to have a certain theoretical level of prediction performance (lower bound on the obtained FIT), based only on SNR level used during experiments:

$$FIT_\gamma < 100 - SNR, \ \gamma = 1, ..., \Gamma - \Phi \quad (9)$$

One is also interested in counting the number of models K that fail to achieve this performance, here called *prediction-wise failures*, and hence defined as models models with a fit value $FIT_\kappa < 70$, $\kappa = 1, ..., K$.

For a control-relevant analysis, a different path may be taken altogether for defining what a failure might be, without the possibility of relating the model performance to the SNR level used for obtaining the estimation data. Computing the maximum, over all the models obtained from the Monte Carlo assessment, of the peak additive uncertainty:

$$\Delta_\gamma(s) = G(s) - \hat{G}_\gamma(s), \ \gamma = 1, ..., \Gamma - \Phi \quad (10)$$

can provide a conservative measure of each algorithm's expected introduced uncertainty for each specific reference system [24]. This uncertainty is quantified in terms of the system-theoretic $\mathcal{H}_\infty$ norm, between the *reference system* and the *worst estimated stochastic model*. The metric shows, in a control sense, the worst-case error expected from each identification algorithm and is reliable provided that a sufficient number of well-designed experiments has been carried.

However, it remains more insightful to assess this uncertainty in a fully statistical sense. Hence, computation of standard statistics (such as mean and standard deviation or median and the associated first and third quartiles) is here opted for, but for a relative measured of uncertainty *i.e.* the *normalized additive uncertainty*, expressed in percentage:

$$\Delta_{norm,\gamma} = 100 \cdot \frac{\|G(s) - \hat{G}_\gamma(s)\|_{H_\infty}}{\|G(s)\|_{H_\infty}}, \ \gamma = 1, ..., \Gamma - \Phi(11)$$

A perfect model will have $\Delta_{norm,\gamma} = 0\%$, while a *control-wise failure* can hence be defined as any obtained numerical model (one for which an estimation failure did not occur) where $\Delta_{norm,\iota} > 100\%$, $\iota = 1, ..., I$ *i.e.* the model's error relative to the system exceeds the true system's $\mathcal{H}_\infty$ norm by a certain factor above unity so the model has more uncertainty than the size of the system; their number I is to be reported as well.

The described reference systems as well as the code used for data generation/processing can be accessed from `http://w3.cran.univ-lorraine.fr/perso/hugues.garnier/Benchmark/Benchmark_Source_Files.rar`.

4 Analysis of Results for Generated Reference Systems

In this section, the described benchmark is used for one Monte Carlo simulation with 100 runs carried for the identification of TF models of the provided reference systems, using the TFEST and SRIVC approaches, for the case when the model orders are the same as the ones of the true system $\mathcal{S}$ *i.e.* $\mathcal{S} \in \mathcal{M}$ where $\mathcal{M}$ denotes the model structure. A related, yet different analysis could be carried when the expected model orders would be different from the true ones. This latter analysis falls outside of the scope of the current study since the goal of the paper has been to provide a benchmark design methodology accompanied by an evaluation of the reliability of two OE model identification algorithms within a standard setup.

Estimating OE model parameters is a non-convex problem, wherein a suitable initialization of the parameter vector is necessary [12]. In general, the algorithms currently investigated within this paper can be very sensitive to the parameter initialization [14]. As there is no

known reliable solution, within the core of the presented study we report both the cases where the initialization is done by the true system's parameters and those when the parameter search is left to its default mode for both algorithms (no specific initialization setting).

Let us first perform such an analysis in a SISO setting. In Figure 4, the boxplots of the obtained FIT values for the 100 Monte Carlo runs are given for both TFEST and SRIVC, with number of prediction-wise failures shown at the top, next to the number estimation failures, shown between round brackets (the circles represent the means). As can be observed, upon algorithm initialization by the true system's parameters, the algorithms perform well with satisfactory overall statistics and medians located around 90%, even though for higher system orders the performance seems to drop. No failure of any kind has been generated throughout the Monte Carlo runs, according to Figure 4 and Figure 5. Note for Figure 5, the statistical drop of relative norm error for TFEST, due to the routine's small numbers of iterations when initialized by the true system's parameters for higher order systems.
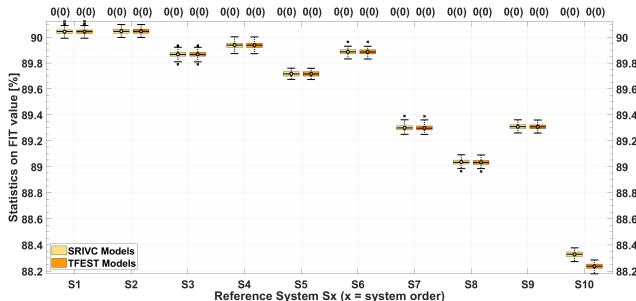

Fig. 4. Boxplots of FIT for 1st to 10th order SISO systems (algorithms on initialization by the true system's parameters). At the top, the number of prediction-wise failures K is given, next to the number of estimation failures Φ shown between round brackets, for each routine. The y-axis is linear but zoomed in a range from 88.2% to 90.2%.
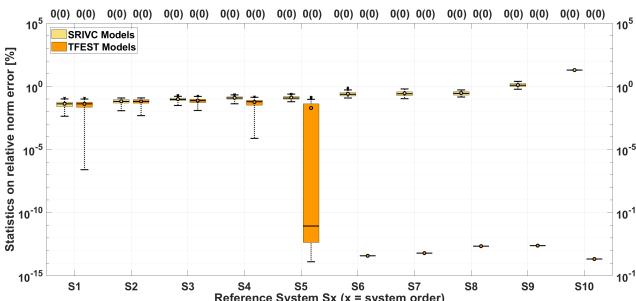

Fig. 5. Boxplots of Δ_{norm} for 1st to 10th order SISO systems (algorithms on initialization by the true system's parameters). At the top, the number of control-wise failures I is given, next to the number of estimation failures Φ shown between round brackets, for each routine. The y-axis is logarithmic.

For default initialization on both algorithms, the fit

statistics are shown in Figure 6. As can be noticed, the medians of the fits continue to be located around 90%, while estimation errors appear present for large order systems for the SRIVC routine and some prediction-wise failures for the TFEST routine for a system of order 10. From a control-theoretic viewpoint, we can see in Figure 7 that the reliability of the two algorithms drops as the system complexity increases, but not considerably for moderately-low system orders (below 5). Nonetheless, for higher system orders, such as for systems of order 9 and 10, based on the medians of the relative norm error defined in equation (11), it can be seen that the generated models can often experience uncertainty of size comparable to the true system's size (norm), indicating that the two methods are less reliable in this sense.
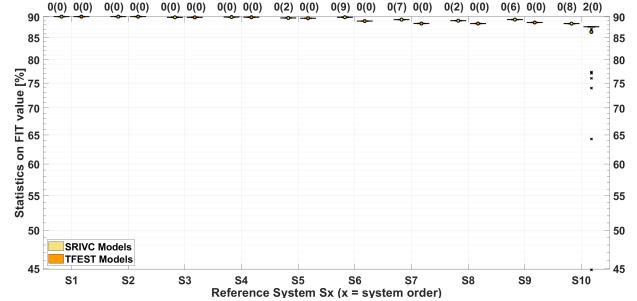

Fig. 6. Boxplots of FIT for 1st to 10th order SISO systems (algorithms on default initialization). At the top, the number of prediction-wise failures K is given, next to the number of estimation failures Φ shown between round brackets, for each routine. The y-axis is linear but zoomed in a range from 45% to 90%.
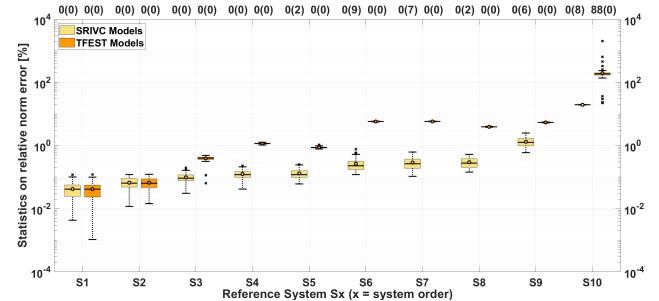

Fig. 7. Boxplots of Δ_{norm} for 1st to 10th order SISO systems (algorithms on default initialization). At the top, the number of control-wise failures I is given, next to the number of estimation failures Φ shown between round brackets, for each routine. The y-axis is logarithmic.

In a MISO setting, for initialization by the true system's parameters on both algorithms, the fit statistics are similarly shown in Figure 8. With medians of the fits still located around 90% for the lower order systems, it can still be statistically noticed that the algorithms' performance slightly drops for larger order systems. Yet, no prediction-wise failures or estimation failures have been detected throughout the 100 Monte Carlo. Nonetheless, with respect to the control-oriented assessment criterion,

we can see that the generated models errors remain low for low order systems and can significantly increase for higher order ones, as shown in Figure 9; a similar conclusion can be drawn based on the numbers of control-wise failures which tend to become large for high order systems (9 or 10) for the models obtained via SRIVC.
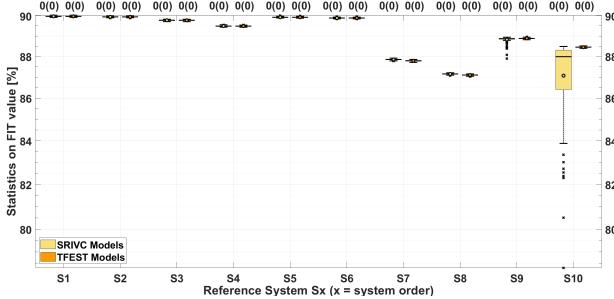

Fig. 8. Boxplots of FIT for 1st to 10th order 2x1 MISO systems (algorithms on initialization by the true system's parameters). At the top, the number of prediction-wise failures K is given, next to the number of estimation failures Φ shown between round brackets, for each routine. The y-axis is linear but zoomed in a range from 78% to 90%.
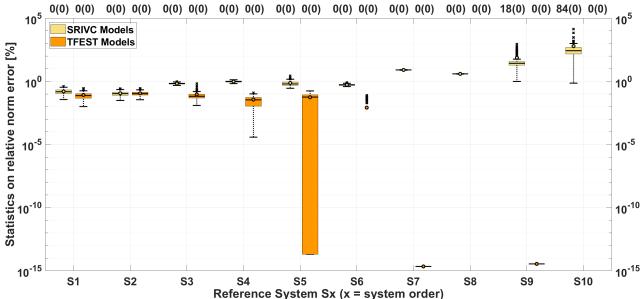

Fig. 9. Boxplots of Δ_{norm} for 1st to 10th order 2x1 MISO systems (algorithms on initialization by the true system's parameters). At the top, the number of control-wise failures I is given, next to the number of estimation failures Φ shown between round brackets, for each routine. The y-axis is logarithmic.

In a similar fashion to the SISO case, the MISO results for default algorithm initialization confirm that the two algorithms are still rather sensitive to the initialization aspect, as confirmed in Figures 10 and 11. The obtained statistics in this case, reported in Figure 10, show that for large system orders the default initialization on the two algorithms renders models of questionable quality, with *e.g.* as many as 34 of the 100 obtained models being prediction-wise failures for the TFEST algorithm (in the case of the 10th order 2x1 MISO system), while as many as 13 of the 100 Monte Carlo runs having generated estimation failures on the SRIVC algorithm (on the same system). From a control-theoretic viewpoint, the statistics on the model errors from Figure 11 confirm the limited algorithms' reliability in providing appropriate models with medians of the relative errors being generally above 90% and many reported control-wise failures for both algorithms. We can hence conclude that even

though theoretical guarantees on the statistical properties of the parameter estimates could otherwise set the standards for the expected potential of identification algorithms [8], implementation plays as large a role on the achieved performance as does a well-designed experiment in practice, especially with respect to the robustness of initialization [22] for multi-input identification.
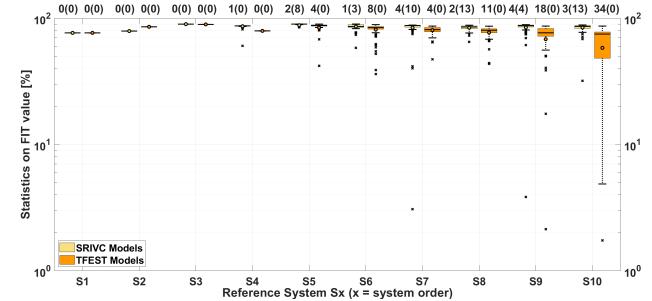

Fig. 10. Boxplots of FIT for 1st to 10th order 2x1 MISO systems (algorithms on default initialization). At the top, the number of prediction-wise failures K is given, next to the number of estimation failures Φ shown between round brackets, for each routine. The y-axis is logarithmic.

Fig. 11. Boxplots of Δ_{norm} for 1st to 10th order 2x1 MISO systems (algorithms on default initialization). At the top, the number of control-wise failures I is given, next to the number of estimation failures Φ shown between round brackets, for each routine. The y-axis is logarithmic.
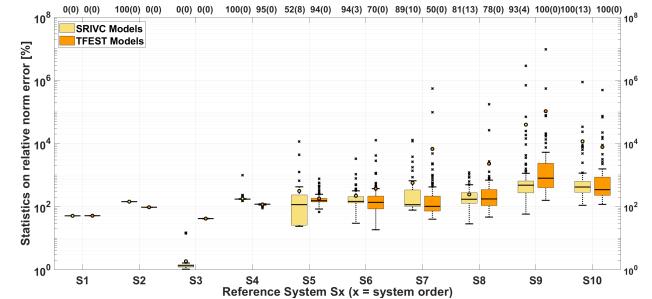
5 Conclusions

In this paper, a study has been carried for numerically assessing the performance of continuous-time model identification methods using an evaluation setup based on computer simulations. The design of the benchmark has been carried according to pragmatic criteria pertaining to randomized system generation; it has been extensively explained and motivated, going through generating reference systems, creating informative orthogonal multisine inputs and choosing appropriate model quality assessment criteria for both a system identification and an automatic control perspective on the obtained results. The main goal of the study has been to use the proposed benchmark towards revealing the main aspects for further consideration in continuous-time model identification; as such, one topic for further research and development has been recognized to be that of algorithm initial-

ization robustness for multi-input linear model identification. The conclusion, though known in theory, has yet to be consistently verified in practice on a test bench. Though the study has been limited in terms of the investigated methods only to the maximum likelihood and the instrumental variable methods for transfer function model identification, the benchmark can be used for similar analyses of other methods. Furthermore, it is hoped that the study itself serves as an encouragement for the system identification community to further use benchmark tests in their investigations.

References

[1] J. D. Alvarez, J. L. Guzmán, D. E. Rivera, M. Berenguel, and S. Dormido. Perspectives on control-relevant identification through the use of interactive tools. *Control Engineering Practice*, 21(2):171–183, February 2013.

[2] T. Chen, H. Ohlsson, and L. Ljung. On the estimation of transfer functions, regularizations and gaussian processes - revisited. *Automatica*, 48(8):1525–1535, June 2012.

[3] T. Dobrowiecki, J. Schoukens, and P. Guillaume. Optimized excitation signals for MIMO frequency response function measurements. In *IMTC 2005 - Instrumentation and Measurement Technology Conference*. Ottawa, Canada, May 2005.

[4] H. Garnier and M. Gilson. CONTSID: a MATLAB toolbox for standard and advanced system identification. In *18th IFAC Symposium on System Identification*. Stockholm, Sweden, July 2018.

[5] H. Garnier, M. Gilson, P. C. Young, and E. Huselstein. An optimal IV technique for identifying continuous-time transfer function model of multiple-input systems. *Control Engineering Practice*, 15(4):471–486, April 2007.

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

[7] M. Gevers, A. S. Bazanella, X. Bombois, and L. Mišković. Identification and the information matrix: how to get just sufficiently rich? *IEEE Transactions on Automatic Control*, 54(12):2828–2840, December 2009.

[8] M. Gevers, L. Mišković, D. Bonvin, and A. Karimi. Identification of multi-input systems: variance analysis and input design issues. *Automatica*, 42(4):559–572, April 2006.

[9] K. Glover. All optimal Hankel-norm approximations of linear multivariable systems and their $\mathcal{L}_\infty$-error bounds. *International Journal of Control*, 39(6):1115–1193, June 1984.

[10] H. Hjalmarsson, C. R. Rojas, and D. E. Rivera. System identification: A Wiener-Hammerstein benchmark. *Control Engineering Practice*, 20(11):1095–1096, November 2012.

[11] A. Kroll and H. Schulte. Benchmark problems for nonlinear system identification and control using soft computing methods: need and overview. *Applied Soft Computing*, 25(12):496–513, December 2014.

[12] L. Ljung. *System Identification - Theory for the User*. Prentice-Hall, New Jersey, U.S.A., 2nd edition, 1999.

[13] L. Ljung. *System Identification Toolbox: User's Guide*. The MathWorks Inc., Natick, Massachussetts, 10th edition, 2018.

[14] L. Ljung. Initialization aspects for subspace and output-error identification methods. In *European Control Conference*. Cambridge, United Kingdom, September 2003.

[15] L. Ljung and R. Singh. Version 8 of the MATLAB System Identification Toolbox. In *16th IFAC Symposium on System Identification*. Brussels, Belgium, July 2012.

[16] L. Ljung and A. Wills. Issues in sampling and estimating continuous-time models with stochastic disturbances. In *IFAC World Congress*. Seoul, Korea, July 2008.

[17] R. Mehra and J.S. Tyler. Case studies in aircraft parameter identification. In *3rd IFAC Symposium on System Identification*. The Hague, The Netherlands, 1973.

[18] M. Opmeer. Decay of Hankel singular values of analytic control systems. *Systems & Control Letters*, 59(10):635–638, October 2010.

[19] V. Pascu, H. Garnier, L. Ljung, and A. Janot. Developments towards formalizing a benchmark for continuous-time model identification. In *55th IEEE Conference on Decision and Control*. Las Vegas, U.S.A., December 2016.

[20] R. Pintelon and J. Schoukens. *System Identification - A Frequency Domain Approach*. IEEE Press, New Jersey, U.S.A., 2nd edition, 2012.

[21] C. R. Rojas, P. E. Valenzuela, and R. A. Rojas. A critical view on benchmarks based on randomly generated systems. In *17th IFAC Symposium on System Identification*. Beijing, China, October 2015.

[22] Y. Rolain and R. Pintelon. Generating robust starting values for frequency-domain transfer function model estimation. *Automatica*, 35(5):965–972, May 1999.

[23] J. Schoukens, J. Suykens, and L. Ljung. Wiener-Hammerstein benchmark. In *15th IFAC Symposium on System Identification*. July 2016, St. Malo, France. As reported on: `https://tc.ifac-control.org/1/1/Data%20Repository/sysid-2009-wiener-hammerstein-benchmark`.

[24] S. Skogestad and I. Postlethwaite. *Multivariable Feedback Control: Analysis and Design*. John Wiley & Sons, Chicester, United Kingdom, 2nd edition, 2005.

[25] R. S. Smith and G. E. Dullerud. Continuous-time control model validation using finite experimental data. *IEEE Transactions on Automatic Control*, 41(8):1094–1105, August 1996.

[26] T. Söderström and P. Stoica. *Instrumental Variable Methods for System Identification*. Lecture Notes in Control and Information Sciences. Springer-Verlag, Berlin, Germany, 1st edition, 1983.

[27] P. C. Young. *Recursive Estimation and Time-Series Analysis. An Introduction for the Student and Practitioner*. Springer-Verlag, Berlin, Germany, 2nd edition, 2011.

[28] P. C. Young, H. Garnier, and M. Gilson. Refined instrumental variable identification of continuous-time hybrid Box-Jenkins models. In H. Garnier and L. Wang, editors, *Identification of Continuous-Time Models from Sampled Data*, pages 91–131. Springer-Verlag, London, 2008.

[29] K. Zhou, J. C. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice-Hall, New Jersey, U.S.A., 1st edition, 1996.