2018 European Control Conference (ECC)
June 12-15, 2018. Limassol, Cyprus

Removal of Insufficiently Informative Data to Support System Identification in MISO Processes

David Arengas and Andreas Kroll[1]

Abstract— **Model accuracy can be degraded by using insufficiently informative data for parameter estimation. This effect can be observed in case of model structures where the input transfer function and the noise transfer function share the same denominator. Equation error models such as ARX, ARMAX or ARARX are examples of model structures exhibiting the latter characteristic. Removal of insufficiently informative data for parameter estimation may increase model accuracy as reported for SISO systems. In this paper, a search method to find informative intervals in MISO systems is introduced. The removal criteria are based on comparison between means of a regression matrix associated to a model class and a threshold related to the conditioning of the matrix. These matrix means are used as quality factor to discard insufficiently informative data. The performance of the search method is demonstrated in a case study.**

Index Terms— **Data screening; multivariable systems; system identification; dynamic data mining.**

I. Introduction

Particularly in case of production plants, performing experiments for system identification can be prohibited. However, process data is usually logged and large data sets are available. Using these historical data sets can be considered for system identification instead of performing new experiments. Nevertheless, these data sets are usually predominantly stationary in case of continuously operated plants as the process is kept in the same operating points due to production requirements. Predominantly stationary data is characterized by seldom transient changes in the inputs or references signals. We will consider a step signal whose length after the step change is around five times the settling time of the system as an example of a predominantly stationary excitation signal. This definition is also valid for ramp signals. A transient change is observed in the process output a time after the step change occurs. The measurements can be considered "poorly" informative since they can yield an ill-posed regression matrix due to the close similarities between the observations. The entire data set should not be used for parameter estimation because model accuracy will be degraded. Selection of "informative" sequences from the original data set may support system identification in situations as formerly described.

Several methods for SISO systems that remove insufficiently informative data to support system identification have been reported in the literature. The method reported in [1] introduces a removal criterion based on the smallest singular value retrieved from the regression matrix of a selected model. A removal criterion based on the gradient of the smallest singular value of a regression matrix is used to discard insufficiently informative data. This removal methodology can also be extended to other model structures where the noise is under modeled *e.g.* ARMAX and ARARX polynomials.

A second method can be found in [2]. Informative data is selected according to successive evaluation tests with increasing complexity. First, transient changes in input and output signals are assessed recursively and an activation of both tests triggers an analytical evaluation of the regression matrix of a selected model structure. This method can be considered as model-based since ARX-Laguerre models are assumed to describe the process. The removal criterion is based on the reciprocal of the condition number computed on the corresponding regression matrix.

Literature on search methods for multivariable processes is scarce. A first attempt of a search method to address multivariable systems can be found in [3]. In this contribution a search method to locate and extract informative sequences from Multi-Input Single-Output (MISO) processes is presented. This paper is organized as follows. A motivating example is introduced in Section II to explore the support to system identification by removing insufficiently informative data. Then, the problem statement and theoretical background are presented in Section III and Section IV respectively. Proposed removal criteria and the search method are described in Section V and Section VI. Performance of the search method is analyzed in Section VII. Finally, conclusions and outlook are presented in Section VIII.

II. Introductory Example

Consider the following system consisting of two inputs and one output which is initially at rest:

$$(1-0.5q^{-1})y(t) = \\ q^{-1}u_1(t)+2q^{-1}u_2(t)+(1+0.8q^{-1}+0.3q^{-2})e(t) \quad (1)$$

where $e(t) \overset{i.i.d.}{\sim} \mathcal{N}(0,\sigma_e^2)$. System (1) is excited by an input signal written in vector notation as:

$$\mathbf{u}^N(t) = \begin{cases} [\mathbf{0}\ \mathbf{0}] & 0 \le t < 20 \\ [\mathbf{4}\ \mathbf{0}] & 20 \le t < 50 \\ [\mathbf{4}\ \mathbf{4}] & 50 \le t \le N-1 \end{cases} \quad (2)$$

with $N = 55, 56, \ldots, 250$. The system has two inputs written in vector notation as $\mathbf{u}(t) = [\mathbf{u}_1(t)\ \mathbf{u}_2(t)],\ t = 0, 1, \ldots, N-1$. Each input is represented by a column vector $\mathbf{u}_i(t) = [u_i(0)\ u_i(1)\ \ldots\ u_i(N-1)]^{\mathrm{T}},\ i = 1, 2$.

[1] David Arengas and Andreas Kroll are with the Faculty of Mechanical Engineering, Department of Measurement and Control, University of Kassel, 34125 Kassel, Germany {david.arengas,andreas.kroll}@mrt.uni-kassel.de

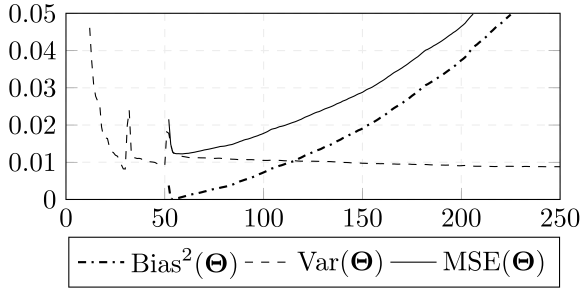

Fig. 1. MSE, squared bias and variance of the parameter vector

The input signal $\mathbf{u}(t)$ is predominantly stationary because its last value is held more than five times the settling time of the system. A multivariable ARX model is selected to identify (1):

$$(1+\theta_1 q^{-1})y(t) = \theta_2 q^{-1}u_1(t) + \theta_3 q^{-1}u_2(t) + \varepsilon(t) \quad (3)$$

The input transfer functions of the system (1) and the model (3) belong to the same model class. In contrast, the noise transfer function of (1) is under-modeled in (3). Monte Carlo simulations were performed on (1) to analyze the behavior of statistical properties of parameters under predominantly stationary signals such as (2). A total of 200 experiments was performed for each length N of the input signal $\mathbf{u}^N(t)$, where each experiment was performed with a different realization of the noise term $e(t)$. The mean squared error of the parameter vector was computed using:

$$\text{MSE}(\hat{\mathbf{\Theta}}(N)) = \sum_{i=1}^{3}\left[\text{Bias}^2(\hat{\theta}_i(N)) + \text{Var}(\hat{\theta}_i(N))\right] \quad (4)$$

Biases, variances and mean squared error (MSE) of the parameter vector as a function of N are depicted in Fig. 1. An increase of the MSE and bias of the parameters is observed at some time after $N = 50$. Moreover, the variance of the parameter vector decays gradually with N.

This simple example shows that using insufficiently informative data degrades model accuracy. Better models can be computed if data is restricted to informative data intervals. The next section aims at introducing formally the problem of removing insufficiently informative data. Most of the notation used in this and further sections is based on [1] and [4].

III. Problem Statement

Consider the following m-inputs one-output system described by a MISO ARMAX polynomial:

$$A_0(q)y(t) = \sum_{i=1}^{m} B_0^{(i)}(q)u_i(t-n_{k_i}+1) + C_0(q)e(t) \quad (5)$$

where $y(t) \in \mathbb{R}$ is the system output, $u_i(t) \in \mathbb{R}$ is the i-th system input and $n_{k_i} \in \mathbb{R}$ is the delay between the system output and the i-th system input $u_i(t)$. The additive noise is described by $e(t) \overset{i.i.d.}{\sim} \mathcal{N}(0,\sigma_e^2)$. The "true" parameters are written in polynomial form as:

$$\begin{aligned}
A_0(q) &= 1 + a_1^0 q^{-1} + \ldots + a_{n_a}^0 q^{-n_a} \quad (6)\\
B_0^{(i)}(q) &= b_{n_{k_i}}^0 q^{-n_{k_i}} + b_{n_{k_i+1}}^0 q^{-n_{k_i}-1} + \ldots \\
&\quad + b_{n_{k_i+n_{b_i}-1}}^0 q^{-n_{k_i}-n_{b_i}+1} \\
i &= 1,2,\ldots,m;\ b_{n_{k_i}}^0 \neq 0 \\
\mathbf{B}_0(q) &= \left[B_0^{(1)}(q)\ B_0^{(2)}(q)\ \ldots\ B_0^{(m)}(q)\right] \quad (7)\\
C_0(q) &= 1 + c_1^0 q^{-1} + \ldots + c_{n_c}^0 q^{-n_c}
\end{aligned}$$

Where $n_{b_i} \in \mathbb{R}^{1\times m}$ is a row vector containing the number of parameters of the B polynomials for each input and n_a and n_c represent the number of parameters of their respective polynomials in (5).

A linear MISO ARX model is used to identify (5):

$$A(q)y(t) = \sum_{i=1}^{m} B^{(i)}(q)u_i(t-n_{k_i}+1) + e(t) \quad (8)$$

The predictor for (8) at any time t is:

$$\hat{y}(t) = \sum_{i=1}^{m} B^{(i)}(q)u_i(t) + (1-A(q))y(t) \quad (9)$$

where each polynomial $B^{(i)}(q)$ contains n_{k_i} delays associated to the dynamic of i-th input to the output. The polynomials of the model are:

$$\begin{aligned}
A(q) &= 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a} \quad (10)\\
B^{(i)}(q) &= b_{n_{k_i}} q^{-n_{k_i}} + b_{n_{k_i+1}} q^{-n_{k_i}-1} + \ldots \\
&\quad + b_{n_{k_i+n_{b_i}-1}} q^{-n_{k_i}-n_{b_i}+1} \\
i &= 1,2,\ldots,m;\ b_{n_{k_i}} \neq 0 \\
\mathbf{B}(q) &= \left[B^{(1)}(q)\ B^{(2)}(q)\ \ldots\ B^{(m)}(q)\right] \quad (11)
\end{aligned}$$

Defining the following arrays,

$$\begin{aligned}
&\boldsymbol{\phi}(t) = [-y(t-1)\ -y(t-2)\ \ldots\ -y(t-n_a) \\
&u_1(t-n_{k_1})\ u_1(t-n_{k_1}-1)\ \ldots\ u_1(t-n_{k_1}+n_{b_1}-1) \\
&u_2(t-n_{k_2})\ u_2(t-n_{k_2}-1)\ \ldots\ u_2(t-n_{k_2}+n_{b_2}-1) \\
&\vdots \\
&u_m(t-n_{k_m})\ u_m(t-n_{k_m}-1)\ u_m(t-n_{k_m}-2)\ \ldots \\
&u_m(t-n_{k_m}+n_{b_m}-1)]^{\mathrm{T}} \quad (12)\\
&\mathbf{\Theta} = \Big[a_1\ a_2\ \ldots\ a_{n_a}\ b_{n_{k_1}}\ b_{n_{k_1+1}}\ \ldots\ b_{n_{k_1+n_{b_1}-1}} \ldots \\
&b_{n_{k_m}}\ b_{n_{k_m+1}}\ \ldots\ b_{n_{k_m+n_{b_m}-1}}\Big]^{\mathrm{T}} \quad (13)
\end{aligned}$$

the process output in (5) can be expressed by:

$$y(t) = \phi^{\mathrm{T}}(t)\mathbf{\Theta}_0 + C_0(q)e(t) \quad (14)$$

with $\mathbf{\Theta}_0$ representing the "true" parameter vector. Analogously, the predictor for (9) is expressed in vector notation using (10) and (11) by:

$$\hat{y}(t,\mathbf{\Theta}) = \phi^{\mathrm{T}}(t)\mathbf{\Theta} \quad (15)$$

where $\mathbf{\Theta}$ represents the model parameters and ϕ is the regression vector. Using matrix notation, (15) is written as:

$$\hat{\mathbf{y}} = \mathbf{\Phi}\mathbf{\Theta} \tag{16}$$

with the definitions,

$$\mathbf{\Phi} = [\phi(1)\ \phi(2)\ \ldots\ \phi(N)]^{\mathrm{T}} \tag{17a}$$

$$\mathbf{y} = [y(1)\ y(2)\ \ldots\ y(N)]^{\mathrm{T}} \tag{17b}$$

The optimal solution for the linear regression in (16) using the method of least squares is

$$\hat{\mathbf{\Theta}}_{LS} = \left(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{y} \tag{18}$$

with $\mathbf{\Phi}$ and $\mathbf{y}$ defined as in (17a) and (17b) respectively. Consider that identification data is generated by exciting the system in (1) with input signals as described by (2). This data set is predominantly stationary since transient changes occur seldom in the excitation signal. The entire data set is represented by $\mathbf{Z}^N = \{y(1), \mathbf{u}(1), y(2), \mathbf{u}(2), \ldots, y(N), \mathbf{u}(N)\}$. Moreover, consider that a subset $\mathbf{Z}^{N_q} \subset \mathbf{Z}^N$ is formed by removing insufficiently informative data from the entire data set:

$$\begin{aligned}\mathbf{Z}^{N_q} = \{&y(k_1), \mathbf{u}(k_1), y(k_1+1), \mathbf{u}(k_1+1), \ldots, \\ &y(k_1^s), \mathbf{u}(k_1^s), y(k_2), \mathbf{u}(k_2), y(k_2+1), \mathbf{u}(k_2+1), \ldots, \\ &y(k_2^s), \mathbf{u}(k_2^s), \ldots, y(k_q), \mathbf{u}(k_q), y(k_q+1), \mathbf{u}(k_q+1), \ldots, \\ &y(k_q^s), \mathbf{u}(k_q^s)\}\end{aligned}$$

with the following conventions:

$$\begin{aligned} k_j^l &: \text{lower bound of an informative interval} \\ k_j^s &: \text{upper bound of an informative interval} \\ \Delta_j &: \text{size of an informative interval} \\ \Delta_j &= k_j^s - k_j^l + 1 \end{aligned}$$

In compact form, $\mathbf{Z}^{N_q}$ can be expressed by:

$$\mathbf{Z}^{N_q} = \left\{\mathbf{Z}\left[k_1^l,\ k_1^s\right],\ \mathbf{Z}\left[k_2^l,\ k_2^s\right], \ldots, \mathbf{Z}\left[k_q^l,\ k_q^s\right]\right\} \tag{19}$$

where $\mathbf{Z}\left[k_j^l,\ k_j^s\right],\ j = 1, 2, \ldots, q$, represents a subset containing informative data extracted from $\mathbf{Z}^N$.

A mapping between the informative subset $\mathbf{Z}^{N_q}$ and the parameter vector can be expressed by:

$$\mathbf{Z}^{N_q} \to \hat{\mathbf{\Theta}}_{N_q} \ \in\ D_{\mathcal{M}} \tag{20}$$

As briefly introduced in Section II, removal of insufficiently informative data may yield more accurate models. Then, the concerned problem can be stated as follows: "Remove insufficiently informative data from a given data set $\mathbf{Z}^N$ aiming at retrieving a subset $\mathbf{Z}^{N_q} \subset \mathbf{Z}^N$ which yields a more accurate model". From Fig. 1, it can be observed that the bias of the estimated input transfer function increases with larger data sets that are predominantly stationary. A model computed with the subset $\mathbf{Z}^{N_q}$ may yield smaller bias than a model computed with the entire data set.

$$\left|\mathbf{G}_0 - \mathbf{G}_{\hat{\mathbf{\Theta}}_{N_q}}\right| < \left|\mathbf{G}_0 - \mathbf{G}_{\hat{\mathbf{\Theta}}_N}\right| \tag{21}$$

Aiming at describing the problem formally, assume that the "true" parameters $\mathbf{\Theta}_0$ are known. Then, if the input transfer functions of the system and model are in the same class and $\mathbf{\Theta}_0$ exists, the problem can be stated as:

$$\begin{aligned}\mathbf{Z}^{N_q} &:= \underset{\mathbf{Z}^{N_q} \subset \mathbf{Z}^N}{\arg\min} \|\hat{\mathbf{\Theta}}(N_q) - \mathbf{\Theta}_0\| \\ &\text{with } \mathbf{Z}^{N_q} = \bigcup_{j=1}^{q} \mathbf{Z}\left[k_j^l,\ k_j^s\right]\end{aligned} \tag{22}$$

Therefore, the aim is to find an "optimal" subset $\mathbf{Z}^{N_q}$, if it exists, which yields the best model. In practice, (22) cannot be implemented since the "true" parameters are not known. A feasible method to select informative sequences will be presented in the next sections.

The models obtained as previously described can also be compared based on the performance of their predictors. Consider the one-step-ahead predictors obtained by two models computed with the entire $\mathbf{Z}^N$ and informative data sets $\mathbf{Z}^{N_q}$. A predictor is obtained from the model $\mathcal{M}(\hat{\mathbf{\Theta}}_{N_q})$ *i.e.* $\hat{\mathbf{y}}(t|t-1, \hat{\mathbf{\Theta}}_{N_q})$. Analogously, the predictor obtained from the model $\mathcal{M}(\hat{\mathbf{\Theta}}_N)$ is $\hat{\mathbf{y}}(t|t-1, \hat{\mathbf{\Theta}}_N)$. The Normalized Mean Squared Error (NMSE) is proposed as comparison criterion. A comparison between the predictors is presented in Section VII.

IV. Theoretical Background

A. Persistence of excitation

Persistence of excitation is a required condition to apply the proposed search method. The input signals treated in this contribution can be considered as "poorly" persistently exciting. Their order of excitation is enough to consistently estimate the number of parameters in the model but they are predominantly stationary. A treatment of persistence of excitation for linear systems is found in [4] and [5]. Moreover, the required order of excitation (or degree of richness [6]) to compute commonly used SISO parametric models in open and closed loop are examined in [6].

Conditions on persistently exciting signals for multivariable systems require a more exhaustive analysis than their counterpart in SISO systems. When a multivariable system is operating in open loop, inputs shall be excited simultaneously [7] aiming at decreasing the variance of the estimated parameters. However, for closed loop operation, it may not be necessary to excite all reference signals due to couplings introduced by the feedback loop [8]. In this contribution, it is assumed that the input signals simultaneously enter into a system operating in open loop which agrees with the analysis in [7].

Definition 1: Persistent excitation−finite-length data MIMO case [9]. A zero mean finite-length stationary process $\mathbf{U} \in \mathbb{R}^{(k_i^s - k_i + 1)\times m}$ is said to be persistently exciting (PE) of order n if the matrix:

$$\mathbf{U}_{N_i - k_i} = \begin{bmatrix} \mathbf{u}(k_i) & \mathbf{u}(k_i+1) & \ldots & \mathbf{u}(k_i^s - k_i - n) \\ \mathbf{u}(k_i+1) & \mathbf{u}(k_i+2) & \ldots & \mathbf{u}(k_i^s - k_i - n + 1) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}(k_i+n) & \mathbf{u}(k_i+n+1) & \ldots & \mathbf{u}(k_i^s - k_i) \end{bmatrix} \tag{23}$$

TABLE I

COMMON MEANS OF MATRIX IN DESIGN OF EXPERIMENTS [10]

Commonly Used Optimality Criteria				
No.	p	Equation	Criterion name	Feature considered
C1	$\gamma_{-\infty}(\mathbf{Q})$	$\lambda_{\min}(\mathbf{\Lambda})$	E-criterion	Smallest-eigenvalue
C2	$\gamma_{-1}(\mathbf{Q})$	$\left(\frac{1}{r}\,\mathrm{trace}\,\mathbf{\Lambda}^p\right)^{1/p}$	A-criterion	Average-variance
C3	$\gamma_0(\mathbf{Q})$	$(\det \mathbf{\Lambda})^{1/r}$	D-criterion	Determinant
C4	$\gamma_1(\mathbf{Q})$	$\frac{1}{r}\,\mathrm{trace}\,\mathbf{\Lambda}^p$	T-criterion	Trace

has rank $n \cdot m$, where $\mathbf{u}(k) = [u_1(k)\ u_2(k)\ \ldots\ u_m(k)]^{\mathrm{T}}$.

The problem stated in (18) can be solved by using SVD on the regression matrix $\mathbf{\Phi}$. An analysis of the eigenvalues associated to each estimated parameter shows the effect of "poorly" PE signals in parameter estimation. In [1], the effect of such signals is evaluated for an ARX SISO model with data collected in open loop. The smallest eigenvalue of the regression matrix increases linearly from an instant which nearly coincides with the instant where the model bias reaches its minimum. Encompassing more identification data further than that instant may degrade the model accuracy. The smallest eigenvalue is considered in design of experiments as a quality measure. Next, relevant means of matrix are introduced and their use for selecting informative sequences is further introduced.

B. Matrix means

Let $\mathbf{I}(\mathbf{\Theta})$ and $\mathbf{Q}(\mathbf{\Theta})$ be symmetric matrices defined by:

$$\mathbf{I}(\mathbf{\Theta}) = (1/\sigma_e^2)\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi} \tag{24}$$

$$\mathbf{Q}(\mathbf{\Theta}) = \sigma_e^2\ \mathbf{I}(\mathbf{\Theta}) = \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi} \tag{25}$$

The regression matrix $\mathbf{\Phi}$ as in (17a) can be split by singular value decomposition (SVD) into:

$$\underset{(N,n_p)}{\mathbf{\Phi}} = \underset{(N,r)}{\mathbf{U}}\ \underset{(r,r)}{\mathbf{\Sigma}}\ \underset{(r,n_p)}{\mathbf{V}^{\mathrm{T}}} \tag{26}$$

where, $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$ and $\sigma_i^2 = \lambda_i\left(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\right)$, $i = 1, 2, \ldots, r$. The eigenvalues of the matrix $\mathbf{Q}$ can be extracted from the diagonal matrix $\mathbf{\Sigma}$ obtained from the singular value decomposition of the regression matrix $\mathbf{\Phi}$. Let $\mathbf{\Lambda}(\mathbf{Q}) = [\lambda_1, \ldots, \lambda_r]^{\mathrm{T}}$ be the vector containing the eigenvalues of $\mathbf{Q}$. The matrix mean γ_p of $\mathbf{Q} \in \mathbb{R}^{r\times r}$ is defined through the vector mean Γ_p which operates over $\mathbf{\Lambda}$ as [10]

$$\gamma_p(\mathbf{Q}) = \Gamma_p(\mathbf{\Lambda}(\mathbf{Q}))\ \forall\ \mathbf{Q} \in \mathrm{Sym}(r),\ p \in [-\infty, \infty] \tag{27}$$

where $\mathrm{Sym}(r)$ represents the space of symmetric matrices of size $r \times r$. Common matrix means used in Design of Experiments are listed in Table I. The E-criterion is used [1] to discard insufficiently informative data. In this contribution, the use of other means of matrix is proposed and evaluated.

V. PROPOSED REMOVAL METHOD

Discarding criteria of insufficiently informative data can be stated based on some of the matrix means in Table I. Let k_s represent the sample when the earliest transient change occurs in one of the input signals of a multivariable system.

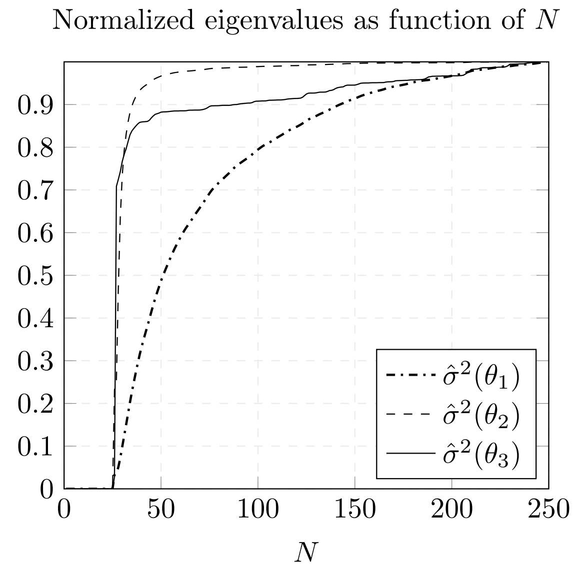

Fig. 2. Normalized eigenvalues associated to the MISO ARX model (3) under an input signal as in (2)

Let $k_j = k_s - n_p$, with n_p the number of parameters, be the starting sample of an informative interval for identification. For a selected model structure, the regression matrix from the instant k_j can be expressed by:

$$\mathbf{\Phi} = [\phi(k_j)\ \phi(k_j+1)\ \ldots\ \phi(N_i)]^{\mathrm{T}} \tag{28}$$

Encompassing observations ϕ_j's that belong to predominantly stationary regions can degrade the model accuracy. An information measure of $\mathbf{\Phi}$ can be stated in terms of one of the matrix means listed in Table I. Therefore, a removal criterion based on a matrix mean of the regression matrix can be formulated as: "discard observations for which the following inequality is satisfied":

$$\begin{aligned}\Gamma_p(\mathbf{\Lambda}(\mathbf{Q}(k^s))) - \Gamma_p(\mathbf{\Lambda}(\mathbf{Q}(k^s-1))) < \eta_c \\ \Gamma_p(\mathbf{\Sigma}^2(k^s)) - \Gamma_p(\mathbf{\Sigma}^2(k^s-1)) < \eta_c\end{aligned} \tag{29}$$

where $\mathbf{Q}(k^s)$ is the symmetric matrix as in (25) evaluated with informative intervals as defined in (19) and $\mathbf{\Sigma}^2(k^s)$ is a vector containing the singular values of $\mathbf{Q}(k^s)$. Fig. 2 depicts the behavior of the singular values associated to each parameter in (3). They behave differently when the system is excited with (2). The smallest eigenvalue, associated to θ_3, jumps abruptly after a step is entered in $\mathbf{u}_2$ at $k = 50$. From some instant later than $k = 50$, the eigenvalue $\hat{\sigma}^2(\theta_3)$ starts to increase linearly because mainly noise is present as "excitation" signal.

Referring to Fig. 2, an information measure based on the E-(C1) criterion can be used to discard insufficiently informative data as originally presented in [1]. In this contribution, the A-(C2) and D-criteria (C3) are also proposed as removal criteria in (29) to discard insufficiently informative data. In order to apply the proposed removal method based on the inequality in (29) the following requirements must be fulfilled:

1) The input signals are persistently exciting of sufficient order to identify the studied process.
2) Transient changes are entered simultaneously in the input system.
3) The system is operating in open loop.
4) The input transfer functions of the model and the system belongs to the same class *i.e.* $\mathbf{G}(\boldsymbol{\Theta}_0) \in \mathbf{G}(\boldsymbol{\Theta})$.
5) The input and noise transfer function of the selected model cannot be independently parametrized. Indeed a multivariable ARX model is used.
6) The additive noise signal is $e(t) \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_e^2)$.

The first two conditions refer to persistence of excitation on the input signals introduced in Section IV-A. Conditions 4 and 5 are required because the removal criterion in (29) is model based. Firstly, fulfillment of conditions related to the input should be evaluated before evaluating the removal criterion. The next section introduces a search method to remove insufficiently informative data.

VI. Search Method

The methodology found in [1] is based on the gradient of the E-criterion (C1) and designed for SISO systems. The novel method proposed here introduces a more general removal procedure based also in the A- (C2) and D-criteria (C3). Therefore, three removal criteria instead of just one criterion are proposed to select informative data in an optimal sense. Additionally, the methodology is extended to multivariable processes in open loop. A methodology found in [2] proposes the use of the reciprocal of the condition number of the information matrix as removal criterion. However, the optimal selection is not explored on that method.

The method proposed here first evaluates transient changes in the inputs which follows a similar procedure to [2]. However, that concept was extended here for multiple inputs. The proposed search method consists of two main stages. It is depicted in Algorithm 1. At first, occurrence of transient changes are evaluated in the input signals by the evaluation test $\mathcal{T}_1$. The sample where a transient change takes place is assigned to be the lower bound, k_i, of a potential informative interval. Then a simultaneous occurrence of the other input signals is explored aiming at meeting the required condition about simultaneous excitation (condition 1). Evaluation of means on the moment matrix is performed in $\mathcal{T}_2$. The E- and A-criteria were used as matrix means for the removal criterion in (29). The regression matrix is expanded by including new regressors as long as the inequality in (29) is satisfied.

VII. Case Study

Consider the double-input one-output (DISO) system initially introduced in Section II. The "true" parameter vector of the input transfer function, $\mathbf{G}_0$, is $\boldsymbol{\Theta}_0 = [-0.5, 1, 2]^{\mathrm{T}}$. Analogously, the parameter vector of the noise transfer function, $\mathbf{H}_0$, is $\boldsymbol{\vartheta}_0 = [-0.5, 0.8, 0.3]^{\mathrm{T}}$. A linear multivariable ARX model is selected to identify the system with parameter vector $\hat{\boldsymbol{\Theta}}(N_q) = [\theta_1, \theta_2, \theta_3]^{\mathrm{T}}$. The system is excited by a predominantly stationary input signal depicted in Fig. 3. Two scales for the y-axis were used. Values of the input signals

Algorithm 1 Search method for finding informative intervals: MISO systems in open loop

Require: m_r, n_a, n_b, η_c
Ensure: Informative subsets $\mathbf{Z}\begin{bmatrix} N_i \\ k_i \end{bmatrix}$, $i = 1, 2, \ldots, q$

```
for k ← 1 to N do
    Evaluate transient changes in u
    if T_1 is true then
        i ← i + 1; k_i ← k
        Do T_2
        k ← k + 1
        Compute Γ_p(Σ²(k))
        while Γ_p(Σ²(k)) − Γ_p(Σ²(k − 1)) > η_c
        N_i ← k
    end if
end for
```

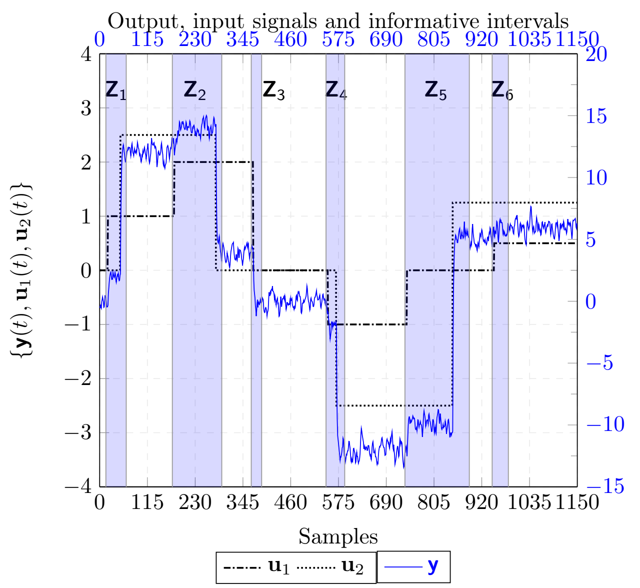

Fig. 3. Excitation signals $\mathbf{u}_1$ and $\mathbf{u}_2$, system response $\mathbf{y}$ and selected and informative intervals $\mathbf{Z}_i$

$\mathbf{u}_1(t)$ and $\mathbf{u}_2(t)$ shall be read from the left y-axis. Moreover, values associated to the output signal $\mathbf{y}(t)$ shall be taken from the right y-axis. Gaussian white noise with $\sigma_e^2 = 0.25$ was added to the output.

It can be noted that the excitation signals are not triggered at the same time index but they act simultaneously from some instant which coincides with the excitation assumption introduced in Section V and with [7]. For instance, consider the second informative interval $\mathbf{Z}_2$ in Fig. 3. An informative interval $\mathbf{Z}_2\,[175,\ 294]$ is located by the search method. The search method notices that u_2 is active but the lower bound of the informative interval $k_2^l = 175$ is marked when u_1 enters into the system. The upper bound is stated according to the proposed removal methodology in (29). A total of 100 experiments was performed on (1). Two parameter vectors were obtained for each experiment: $\hat{\boldsymbol{\Theta}}(N_q)$ and $\hat{\boldsymbol{\Theta}}(N)$. The parameter vector $\hat{\boldsymbol{\Theta}}(N_q)$ was computed using informative

TABLE II

PARAMETER ESTIMATION WITH $\mathbf{Z}^{N_q}$ AND $\mathbf{Z}^N$

Removal criterion		$\mathbf{Z}^{N_q}$ $\gamma_{-\infty}$ C1	$\mathbf{Z}^{N_q}$ γ_{-1} C2	$\mathbf{Z}^{N_2}$ κ_2^{-1}	$\mathbf{Z}^N$
Success rate		96 %	92 %	94 %	–
MSE($\hat{\mathbf{\Theta}}$)		0.0484	0.0442	0.1200	0.3294
Bias^2($\hat{\mathbf{\Theta}}$)		0.0428	0.0389	0.1153	0.3258
Var($\hat{\mathbf{\Theta}}$)		0.0056	0.0054	0.0047	0.0037
NMSE($\hat{\mathbf{y}}, \mathbf{y}$)	Mean	0.0031	0.0032	0.0034	0.0036
	Var $\cdot 10^{-6}$	0.0843	0.1547	0.1602	0.1205

$\mathbf{Z}^{N_q}$: informative subset using the E-(C1) (left column) and A-criteria (C2) (right column). See criteria on Table I.
$\mathbf{Z}^{N_2}$: informative subset using [2].
$\mathbf{Z}^N$: entire data set.

intervals retrieved by the search method. In contrast, $\hat{\mathbf{\Theta}}(N)$ was computed using the entire data set. These vectors were stored for each experiment and their experimental mean was computed to obtain $\tilde{\mathbf{\Theta}}(N_q)$ and $\tilde{\mathbf{\Theta}}(N)$, respectively.

A search method including two matrix means as removal criteria was implemented. The A-(C1) and E-criteria (C2) were used as matrix means in the inequality (29). Table II summarizes the results obtained from the experiments. The performance of the search method was evaluated on the estimated parameters, $\hat{\mathbf{\Theta}}$, and output predictors, $\hat{\mathbf{y}}$. The informative subset and entire data set are represented by $\mathbf{Z}^{N_q}$ and $\mathbf{Z}^N$, respectively. The column for the informative subset is split into three columns. The left column contains the results obtained using the E-criterion (C1). The middle column corresponds to the results of the A-criterion. The right column has the results when a removal criterion based on the reciprocal of the condition number is used as reported in [2]. The depicted success rate refers to the frequency of obtaining more accurate models when they were computed with informative subsets retrieved by the search method. The removal criteria proposed in Section V yield better results than the one proposed in [2]. It can also be noted that the removal criterion based on the smallest eigenvalue (E-criterion) yields a slightly better performance than the method based on the trace of the moment matrix (A-criterion).

Three statistical properties were evaluated for the estimated parameters: bias, variance and mean square error (MSE). The MSE and the bias of the parameters are smaller if the models are computed with informative subsets. This outcome is consistent for the two chosen removal criteria as shown in Table II. In contrast, the parameter variance is smaller for parameters estimated with the entire data set. This can be explained by the asymptotic behavior with the data length N as explained in [1]. However, a smaller parameter variance cannot compensate the reduction in the parameter biases. As a result, the MSE of the parameters is smaller as product of using the informative subset retrieved by the search method.

Output predictors as in (9) were computed in each experiment for four models. Two models are obtained using informative subsets $\mathbf{Z}^{N_q}$ retrieved by the removal criteria proposed in this publication. A third model is computed using the method found in [2]. A fourth model is obtained using the entire data set $\mathbf{Z}^N$. The normalized mean square error (NMSE) between the predictors and the system output are also listed in Table II. The predictors of the models computed with informative subsets yield a smaller NMSE than their counterpart with the entire data set. Simulation results confirm that removal of insufficiently informative data can yield to more accurate models supporting system identification when data are predominantly stationary.

VIII. CONCLUSION AND OUTLOOK

A search method to remove insufficiently informative data from MISO systems for model estimation was presented. It consists of two main evaluation tests: detection of transient changes and evaluation of the gradient of a suitable matrix mean. The E- and A-criteria were implemented in this contribution and it was observed that both yield similar results when they are used as removal criteria. The gradient of the matrix means described by these criteria is compared with a threshold defined by the user and the upper bound of the interval is set when a condition on the gradient is fulfilled. Results show that the proposed search method retrieve subsets that yield more accurate models. As future work further tests on artificial data from a benchmark problem will be performed. Moreover, an extension to systems operating in closed loop is planned. Experiments and tests on real systems are also considered as part of further research.

ACKNOWLEDGEMENT

This work was supported by the Foundation for the Future of Colombia (COLFUTURO) and the German Academic Exchange Service (DAAD).

REFERENCES

[1] P. Carrette, G. Bastin, Y. Y. Genin, and M. Gevers, "Discarding data may help in system identification," *IEEE Transactions on Signal Processing*, vol. 44, no. 9, pp. 2300–2310, 1996.

[2] A. C. Bittencourt, A. J. Isaksson, D. Peretzki, and K. Forsman, "An algorithm for finding process identification intervals from normal operating data," *Processes*, vol. 3, no. 3, pp. 357–383, 2015.

[3] D. Arengas and A. Kroll, "A Search Method for Selecting Informative Data in Predominantly Stationary Historical Records for Multivariable Systems," in *Proceedings of the 21st International Conference on System Theory, Control and Computing ICSTCC 2017*, Sinaia, Romania, October 19-21 2017.

[4] L. Ljung, *System Identification: Theory for the user*, 2nd ed., T. Kailath, Ed. Prentice Hall, 1999.

[5] T. Söderström and P. Stoica, *System Identification*. Prentice Hall International, 1989.

[6] 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*, vol. 54, no. 12, pp. 2828–2840, 2009.

[7] M. Gevers, L. Mišković, D. Bonvin, and A. Karimi, "Identification of multi-input systems: variance analysis and input design issues," *Automatica*, vol. 42, no. 4, pp. 559–572, 2006.

[8] A. S. Bazanella, M. Gevers, and L. Mišković, "Closed-Loop Identification of MIMO Systems: A New Look at Identifiability and Experiment Design," *European Journal of Control*, vol. 16, no. 3, pp. 228–239, 2010.

[9] T. Katayama, *Subspace Methods for System Identification*. Springer, 2005.

[10] F. Pukelsheim, *Optimal Design of Experiments*, R. E. O'Malley, Ed. SIAM, 2006.