# Virtual Reference Feedback Tuning With Data Selection Criteria Applied to a Thermal Process ⋆

**Cristiane Silva Garcia** * **Alexandre Sanfelici Bazanella** *

* *Programa de Pós-Graduação em Engenharia Elétrica (PPGEE), Universidade Federal do Rio Grande do Sul, RS, Brazil, (e-mail: cristianesgcris@gmail.com, bazanella@ufrgs.br).*

**Abstract:** Adjusting the controller's parameters using a data-driven (DD) methodology usually requires data gathered from a specific experiment performed in the process, which may be a time consuming task for the designer. To avoid this task, routine operating data could be used instead. However, combining those raw data along with a DD method almost certainly results in inappropriate tuning. Therefore, it is advisable to pre-select the useful information before estimating the controller's parameters. The present work is an extension of our previous works, where two data selection criteria were applied to the Virtual Reference Feedback Tuning (VRFT) method. In the present work, we have combined the application of those criteria to select the relevant data subsets. Moreover, the controller's parameters are estimated using not only VRFT's original solution, known as Ordinary Least Squares (OLS), but also the Data Least Squares (DLS) solution. The feasibility of the proposed solution is evaluated through experiments carried out in a thermal process.

*Keywords:* Data-driven control, VRFT, data selection criteria, singular values, condition number, errors-in-variables problems, ordinary least squares, data least squares.

## 1. INTRODUCTION

In the data-driven (DD) control framework, a parametrized controller structure is chosen *a priori*, and the controller's parameters are estimated *directly* using data collected from the process, without knowing the model of this process. The quality of the estimate depends on the informativity of the data collected, therefore, it is a common practice to collect data from a specific experiment where a sufficiently rich signal excites the process' input (Bazanella et al., 2011). However, in some cases, performing this specific experiment can be a very costly and even undesirable task (Bittencourt et al., 2015; Shardt and Huang, 2013).

Aiming to avoid this task, one could use routine operating data to adjust the controller's parameters, but, applying such data along with a DD methodology may lead to inappropriate tuning, due to their low informativeness and the presence of noise. However, this does not mean that routine operating data does not have any useful information to tune the controller's parameters. In fact, it is not unusual the presence of relevant intervals inside those data that, if correctly identified, may result in a satisfactory estimate.

The task of searching for relevant subsets is a matter already studied in the system identification framework. As can be seen through several works in the literature (Gevers et al., 2009; Carrette et al., 1996; Bittencourt et al., 2015; Arengas and Kroll, 2017; Shardt and Huang, 2013). However, the task of searching for useful data subsets is a subject that has received little attention in the DD framework. In our previous works, we have adapted two data selection criteria, originally developed for system identification, to be used with the Virtual Reference Feedback Tuning (VRFT) method (Garcia and Bazanella, 2019, 2020). Garcia and Bazanella (2019) approached the case of full-order controllers, that is, the ideal controller may be constructed using the available controller structure. The reduced-order controllers case was considered in Garcia and Bazanella (2020), where we have shown that only one of the criteria could be applied to this specific case. However, in those works, we have separately applied the criteria and considered only the original solution for the VRFT method, known as Ordinary Least Squares (OLS). Moreover, the feasibility of the proposed solution was presented through Monte Carlo simulations.

In the present work, we have combined the application of those two criteria to select the relevant data subsets. Another contribution of the present work is to apply a different solution for the VRFT method. This solution is known as Data Least Squares (DLS) and treats the case when the noise contributions are in the controller's virtual input, which suits well to the VRFT problem when the data are collected during an open-loop experiment. Finally, the applicability of the proposed solution is evaluated through experiments performed in a thermal process.

The remaining of the paper is organized as follows. Section 2 presents the preliminaries. The VRFT method and the OLS and DLS solutions are shown in section 3. Section 4 introduces the data selection criteria. The experimental results are presented in section 5. Finally, section 6 presents the conclusions and future work.

⋆ This work was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior — Brasil (CAPES) — Finance Code 001. This work was financed in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

## 2. PRELIMINARIES

### *2.1 Background*

Consider a linear discrete-time single-variable process

$$y(t) = G(q)u(t) + H(q)v(t), \tag{1}$$

where $q$ is the forward-shift operator, $G(q)$ is the process transfer function, $H(q)$ is the noise model, whereas $y(t)$ is the output signal, $u(t)$ is the input signal, and $v(t)$ is a zero-mean white noise with variance $\sigma_v^2$. It is assumed that $G(q)$ and $H(q)$ are rational and causal transfer functions.

The process (1) is controlled by a linear time-invariant controller with a fixed transfer function. The control action, $u(t)$, is given by

$$u(t) = C(q,\rho)\left[r(t) - y(t)\right], \tag{2}$$

where $C(q,\rho)$ is the controller's transfer function, which is parametrized by a parameter vector $\rho \in \mathbf{R}^m$, where $m$ is the number of parameters, whereas, $r(t)$ is the reference signal. Is is assumed that $r(t)$ is a quasi-stationary signal and uncorrelated with the noise, that is, $\overline{\mathrm{E}}\left[r(t)v(t-\tau)\right] = 0\ \forall\tau$, and $\overline{\mathrm{E}}\left[f(t)\right] \triangleq \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \mathrm{E}\left[f(t)\right]$ with $\mathrm{E}\left[\cdot\right]$ denoting expectation. When collecting data from a closed-loop experiment the input-output signals are

$$y(t) = T(q,\rho)r(t) + S(q,\rho)H(q)v(t), \tag{3}$$

$$u(t) = S(q,\rho)C(q,\rho)\left[r(t) - H(q)v(t)\right], \tag{4}$$

$$S(q,\rho) = \left[1 + C(q,\rho)G(q)\right]^{-1}, \tag{5}$$

$$T(q,\rho) = S(q,\rho)G(q)C(q,\rho), \tag{6}$$

where (3) is obtained replacing (2) in (1), and (4) is obtained replacing (1) in (2). Moreover, $S(q,\rho)$ and $T(q,\rho)$ are the sensitivity function and the complementary sensitivity function, obtained with the controller $C(q,\rho)$, respectively. On the other hand, when collecting data during an open-loop experiment $G(q)$ is assumed to be bounded-input bounded-output (BIBO) stable, $u(t)$ is an exogenous signal, and the output signal is given by (1).

### *2.2 Model reference control*

In the model reference control, the desired closed-loop behaviour is defined *a priori*, through the reference model, $T_d(q)$. The objective is to find the optimal parameter vector $\rho^\star_{\mathrm{mr}}$ that minimizes, for a specified reference signal, the following cost function

$$\rho^\star_{\mathrm{mr}} = \arg\min_{\rho} J^{\mathrm{mr}}(\rho), \tag{7}$$

$$\begin{aligned} J^{\mathrm{mr}}(\rho) &\triangleq \overline{\mathrm{E}}\left[y(t,\rho) - y_d(t)\right]^2 \\ &= \overline{\mathrm{E}}\left[\left(T(q,\rho) - T_d(q)\right)r(t)\right]^2, \end{aligned} \tag{8}$$

where $y_d(t) = T_d(q)r(t)$ is the desired closed-loop response, and $y(t,\rho) = T(q,\rho)r(t)$ is the closed-loop response obtained with the controller $C(q,\rho)$.

The controller parametrization delimits a subset within the controller set, known as the *controller class*, defined as $\mathcal{C} = \{C(q,\rho) | \rho \in \Omega \subseteq \mathbf{R}^m\}$, where $\Omega$ is the subset of all implementable parameter vectors. From (8) and considering (6), the *ideal controller* could be calculated as

$$C_d(q) = \left[G(q) - G(q)T_d(q)\right]^{-1} T_d(q).$$

Considering that the ideal controller is in the controller's class $\mathcal{C}$, i. e., the full-order controllers case, one may have the following assumption.

*Assumption 2.1.* (Matching condition).

$$\exists\, \rho_d \in \Omega \,|\, C(q,\rho_d) = C_d(q),$$

where $\rho_d$ is the *ideal parameter vector.*

A very common choice for the controller's structure is a linearly parametrized (LP) controller as described by Definition 2.1.

*Definition 2.1.* (Linear parametrization). The controller's transfer function can be written as $C(q,\rho) = \rho^{\mathrm{T}}\overline{C}(q)$, where $\overline{C}(q)$ is a vector of $m$ rational transfer functions independent of $\rho$.

## 3. VIRTUAL REFERENCE FEEDBACK TUNING

VRFT is based on the model reference paradigm, this way, the controller's parameters are estimated using the reference model and the controller structure, both chosen *a priori*, and a batch of input-output data collected from the process (Campi et al., 2002; Bazanella et al., 2011).

Those informations are used to generate the signals of the VRFT's *virtual experiment.* From the output data collected the *virtual reference* signal is generated as $\overline{r}(t) = T_d^{-1}(q)y(t)$, which is the signal that should be applied to the desired closed-loop to generate the output signal collected. Using $\overline{r}(t)$, the *virtual error* is calculated as $\overline{e}(t) = \left[T_d^{-1}(q) - 1\right]y(t)$. Considering a LP controller, Definition 2.1, and the virtual error, $\overline{e}(t)$, the regressor vector $\varphi(t)$ is defined as

$$\varphi(t) = \overline{C}(q)\left[T_d^{-1}(q) - 1\right]y(t), \tag{9}$$

which is the input of the ideal controller.

From the above information, the VRFT recasts the problem of minimizing $J^{\mathrm{mr}}(\rho)$, (8), into a least squares identification of the controller

$$J^{\mathrm{vr}}(\rho) = \overline{\mathrm{E}}\left[u_L(t) - \rho^{\mathrm{T}}\varphi_L(t)\right]^2, \tag{10}$$

where

$$\varphi_L(t) = L(q)\varphi(t) \tag{11}$$

$$u_L(t) = L(q)u(t) \tag{12}$$

and $L(q)$ is a filter whose structure depends on the problem to be considered. The minimum of $J^{\mathrm{vr}}(\rho)$ is proven to coincide with the minimum of $J^{\mathrm{mr}}(\rho)$ when it is assumed that the process is not affected by noise (meaning $v(t) = 0$ in (3) and (4), for the closed-loop case, or in (1), for the open-loop case), and that there is an ideal controller $C_d(q)$ such that $J^{\mathrm{mr}}(\rho_d) = 0$, that is, $T(q,\rho_d) = T_d(q)$.

### *3.1 Solutions for the VRFT problem*

A general formulation for the problem of estimating the parameter vector $\rho$ is

$$\left(\Phi_L - \Delta_\Phi\right)\rho = u_L - \delta_u, \tag{13}$$

where the regressor matrix $\Phi_L \in \mathbf{R}^{N\times m}$ and the output vector $u_L \in \mathbf{R}^N$ are given by

$$\Phi_L = \left[\varphi_L(1)\ \varphi_L(2)\ \ldots\ \varphi_L(N)\right]^{\mathrm{T}}, \tag{14}$$

$$u_L = \left[u_L(1)\ u_L(2)\ \ldots\ u_L(N)\right]^{\mathrm{T}}, \tag{15}$$

whereas $\Delta_\Phi \in \mathbf{R}^{N\times m}$ and $\delta_u \in \mathbf{R}^N$ represent the noise contributions in $\Phi_L$ and $u_L$, respectively, and $N$ is the number of samples. Notice that, $\Phi_L = \Phi_0 + \Delta_\Phi$ and

$u_L = u_0 + \delta_u$ where $\Phi_0 \in \mathbf{R}^{N\times m}$ is the noise-free regressor matrix and $u_0 \in \mathbf{R}^N$ is the noise-free output vector, given by

$$\Phi_0 = [\varphi_0(1)\ \varphi_0(2)\ \ldots\ \varphi_0(N)]^{\mathrm{T}},$$
$$u_0 = [u_0(1)\ u_0(2)\ \ldots\ u_0(N)]^{\mathrm{T}}.$$

Therefore, in the noise-free case the problem is $\Phi_0\rho_d = u_0$, that is, $\Delta_\Phi = 0$ and $\delta_u = 0$.

### *3.2 Original solution*

The estimated parameters are obtained by minimizing the squares of the difference between $\rho^{\mathrm{T}}\varphi_L(t)$ and $u_L(t)$, (10), solving the following normal equation:

$$\hat{\rho} = \left[\sum\nolimits_{t=1}^{N} \varphi_L(t)\varphi_L^{\mathrm{T}}(t)\right]^{-1} \sum\nolimits_{t=1}^{N} \varphi_L(t)u_L(t), \qquad (16)$$

where $N$ is the number of samples collected in the experiment. In vector form, (16) is rewritten as

$$\hat{\rho} = \left(\Phi_L^{\mathrm{T}}\Phi_L\right)^{-1}\Phi_L^{\mathrm{T}}u_L, \qquad (17)$$

where the regressor matriz $\Phi_L$ and the output vector $u_L$ are given by (14) and (15), respectively.

This solution is known in the literature as OLS and it gives a unbiased estimate under the assumption that only $u_L$ is affected by noise (Van Huffel and Lemmerling, 2013). The OLS problem is formulated setting $\Delta_\Phi = 0$ in (13) as

$$\min_{\delta_u,\rho} \|\delta_u\|_2^2 \quad \text{s. t. } \Phi_L\rho = u_L - \delta_u.$$

However, the VRFT problem is not a standard identification problem, because the virtual input of the controller to be identified is always affected by noise, see (1), while the virtual output is affected by noise only when closed-loop data is used, see (3) and (4). Therefore, the OLS solution gives a biased estimate in the presence of noisy data.

An instrumental variable (IV) method is usually employed to reduce the bias of the estimate. However, the decrease in the bias comes with the cost of increasing its variance (Bazanella et al., 2011). Because of that, the IV approach is not used in the present work.

### *3.3 Data Least Squares solution*

The DLS solution is formulated considering that the regressor matrix is affected by noise and the output vector is known exactly, that is, $u_L$ is not corrupted by noise (DeGroat and Dowling, 1993). The DLS problem is formulated, setting $\delta_u = 0$ in (13), as

$$\min_{\Delta_\Phi,\rho} \|\Delta_\Phi\|_F^2 \quad \text{s. t. } (\Phi_L - \Delta_\Phi)\,\rho = u_L,$$

where $\|\cdot\|_F$ is the Frobenius norm, that is, $\|X\|_F^2 = \sum_{i,j}|x_{i,j}|^2$. The solution for this problem is given in Theorem 3.1.

*Theorem 3.1.* (from DeGroat and Dowling (1993)). The solution for the DLS problem is given by

$$\hat{\rho} = \frac{u_L^{\mathrm{T}}u_L}{u_L^{\mathrm{T}}\Phi_L\underline{v}}\underline{v},$$

where $\underline{v}$ is the smallest right singular vector of $P_b^{\perp}\Phi_L$, and $P_b^{\perp} \in \mathbf{R}^{N\times N} = \left[I - u_L\left(u_L^{\mathrm{T}}u_L\right)^{-1}u_L^{\mathrm{T}}\right]$ is a projection matrix that projects the column space of $\Phi_L$ into the orthogonal complement of $u_L$.

## 4. DATA SELECTION CRITERIA

The objective of this section is to briefly present the data selection criteria applied to the VRFT problem. Subsection 4.1 apresents the Smallest Singular Value (SSV) criterion, the Condition Number (CN) criterion is presented in Subsection 4.2, and the combination of those two criteria is presented in Subsection 4.3.

### *4.1 SSV criterion*

This criterion was adapted to the VRFT problem as presented in Garcia and Bazanella (2019, 2020), and originally developed to the system identification framework in Carrette et al. (1996). The criterion's main idea relies on the fact that when the regressor matrix $\Phi_L$ presents a singular value significantly smaller than the others, the associate eigenparameter is the most affected by noise, therefore, it is enough to evaluate only the SSV of the regressor matrix.

Consider the square of the SSV of $\Phi_L(t)$ given by $\underline{\sigma}^2\left(\Phi_L(t)\right) = \underline{\lambda}\left(P(t)\right)$, where $\underline{\sigma}^2$ is the square of the smallest singular value of the regressor matrix, calculated up to the $t$-th data sample, $\underline{\lambda}(P(t))$ is the smallest eigenvalue of the information matrix $P(t)$, defined as

$$P(t) = \Phi_L^{\mathrm{T}}(t)\Phi_L(t), \qquad (18)$$

also calculated up to the $t$-th data sample, and $t = m, \ldots, N$ to ensure that the information matriz will be non-singular.

This criterion is based on tracking the evolution of the first difference of $\underline{\sigma}^2\left(\Phi_L(t)\right)$. This quantity is defined as

$$\Delta\underline{\sigma}^2\left(\Phi_L(t)\right) = \underline{\sigma}^2\left(\Phi_L(t)\right) - \underline{\sigma}^2\left(\Phi_L(t-1)\right), \qquad (19)$$

where $\Delta\underline{\sigma}^2\left(\Phi_L(t)\right)$ is the first difference of $\underline{\sigma}^2\left(\Phi_L(t)\right)$ and $t$ is the sample index. The criterion is

$$\Delta\underline{\sigma}^2\left(\Phi_L(t)\right) > \eta_{\text{ssv}}, \qquad (20)$$

where $\eta_{\text{ssv}}$ is a threshold chosen by the designer. How to choose a reasonable value for this threshold will be presented in Section 5. Therefore, the regressor vector $\varphi_L(t)$ and the output sample $u_L(t)$ are maintained whenever the inequality in (20) is satisfied. The SSV criterion is summarized in Algorithm 1. The regressor matrix $\Phi_{\text{ssv}}$ and the output vector $u_{\text{ssv}}$ selected with the SSV criterion are employed to estimate the controller's parameters using the OLS solution, (17), and the DLS solution in Theorem 3.1.

### *4.2 CN criterion*

The condition number is calculated as

$$\kappa\left(P(t)\right) = \frac{\overline{\sigma}\left(P(t)\right)}{\underline{\sigma}\left(P(t)\right)}, \qquad (21)$$

where $\overline{\sigma}\left(P(t)\right)$ and $\underline{\sigma}\left(P(t)\right)$ are the largest and smallest singular values of $P(t)$, respectively.

This criterion considers that as the input signal $u(t)$ is little informative and its contributions dominate over the noise, the information matrix $P(t)$ becomes ill-conditioned with time. Therefore, the idea is that reducing the number of samples may prevent the degradation of the condition number. This data selection criterion is applied separately for each step change, and consists in maintaining the

ISSN: 2175-8905

DOI: 10.20906/sbai.v1i1.2759

**Algorithm 1:** SSV criterion

**Data:** Reference model $T_d(q)$, vector $\overline{C}(q)$, filter $L(q)$, threshold $\eta_{\text{ssv}}$, number of paramters to be estimated $m$, input vector $u$ and output vector $y$.
**Result:** Regressor matrix $\Phi_{\text{ssv}}$ and output vector $u_{\text{ssv}}$ selected with the criterion.

**calculate** $\Phi_L$ as in (14) using (11) and (9) ;
**calculate** $u_L$ as in (15) using (12) ;
$\Phi_{\text{ssv}} \leftarrow 0$ ;
$u_{\text{ssv}} \leftarrow 0$ ;
$t \leftarrow m+1$ ;
**while** $t < N$ **do**
  **calculate** $\Delta\underline{\sigma}^2\left(\Phi_L(t)\right)$ as in (19) ;
  **if** $\Delta\underline{\sigma}^2\left(\Phi_L(t)\right) > \eta_{ssv}$ **then**
    $\Phi_{\text{ssv}} \leftarrow [\Phi_{\text{ssv}} \;\; \varphi_L(t)]$ ;
    $u_{\text{ssv}} \leftarrow [u_{\text{ssv}} \;\; u_L(t)]$ ;
  **end**
  $t \leftarrow t+1$ ;
  **next iteration**;
**end**

regressor rows $\varphi_L(t)$, and the corresponding output rows $u_L(t)$, while the following inequality holds:

$$\kappa\left(P(t)\right) > \eta_{\text{cn}}, \tag{22}$$

where $\eta_{\text{cn}}$ is a threshold chosen by the designer, and the sample index $t$ varies from the start of the transient until the end of the interval. This way, the informative subset is delimited by the start of the transient caused by an input change, up to, but not including, the sample when the inequality (22) stops holding. Algorithm 2 summarizes the CN criterion. The regressor matrix $\Phi_{\text{cn}}$ and the output vector $u_{\text{cn}}$ selected with the CN criterion are employed to estimate the controller's parameters using the OLS solution (17) and the DLS solution in Theorem 3.1.

### *4.3 Combining the criteria*

The combination of the criteria is accomplished this way: the lines of $\Phi_{\text{ssv}}$ are maintained if they also correspond to lines of $\Phi_{\text{cn}}$. Equivalently, the lines of $u_{\text{ssv}}$ are maintained if they correspond to lines of $u_{\text{cn}}$. The controller's parameter vector is estimated using the OLS and DLS solutions and the regressor matrix and output vector selected combining the criteria.

**Algorithm 2:** CN criterion

**Data:** Reference model $T_d(q)$, vector $\overline{C}(q)$, filter $L(q)$, threshold $\eta_{\text{cn}}$, number of parameters to be estimated $m$, input vector $u$ and output vector $y$.
**Result:** Regressor matrix $\Phi_{\text{cn}}$ and output vector $u_{\text{cn}}$ selected with the criterion.

$\Phi_{\text{cn}} \leftarrow 0$ ;
$u_{\text{cn}} \leftarrow 0$ ;
$t \leftarrow m$ ;
**calculate** $\Phi_L$ as in (14) using (11) and (9) ;
**calculate** $u_L$ as in (15) using (12) ;
**while** $t \leq N$ **do**
  **calculate** $P(t)$ as in (18) ;
  **calculate** $\kappa\left(P(t)\right)$ as in (21) ;
  **if** $\kappa\left(P(t)\right) \geq \eta_{cn}$ **then** **execution ending**;
  $\Phi_{\text{cn}} \leftarrow [\Phi_{\text{cn}} \;\; \varphi_L(t)]$ ;
  $u_{\text{cn}} \leftarrow [u_{\text{cn}} \;\; u_L(t)]$ ;
  $t \leftarrow t+1$ ;
  **next iteration**;
**end**

## 5. EXPERIMENTAL RESULTS

In order to evaluate the application of the data selection criteria, presented in the previous section, an experiment in a thermal process is executed to collect data. This thermal process is composed by a heating element, a temperature sensor, and a commercial Proportional-Integral-Derivative (PID) controller (Novus, 2021). The heating element is a resistor that dissipates heat by the Joule effect, and the sensor is a type K thermocouple that converts temperature into a voltage difference. The voltage on the heater is controlled by a pulse width modulated (PWM) signal, and the process' input is the signal pulse width percentage. The process' output is the voltage on the thermocouple converted into a temperature in Celsius degrees.

To estimate the controller's parameters, first, the data are collected from an open-loop experiment where the input signal is composed by two steps with PWM amplitudes of 7% and 8%, respectively. Each step comprises 1800 samples. Figure 1 presents the input $u(t)$ and the output $y(t)$ signals collected during the experiment. Observe in this figure that there are measurement errors in the output signal.

The open-loop settling time is approximately 500 samples, inferred from Figure 1. Observe that, this process can be approximated by a first-order transfer function. Therefore, by choosing a desired closed-loop behaviour faster than open-loop and using a controller with a PI structure, it is expected a resulting behaviour with overshoot. Because of that, the reference model was chosen with a complex-conjugate pole pair, as given by

$$T_d(q) = \frac{0.051064(q-0.9546)}{(q^2 - 1.94q + 0.9419)}, \tag{23}$$

providing settling time of 100 samples. The controller to be estimated is a PI with the following structure

$$C(q,\rho) = [K_p \;\; K_i]\begin{bmatrix}1 & \dfrac{q}{q-1}\end{bmatrix}^{\mathrm{T}}, \tag{24}$$

where $\rho = [K_p \;\; K_i]^{\mathrm{T}}$ is the parameter vector to be estimated.

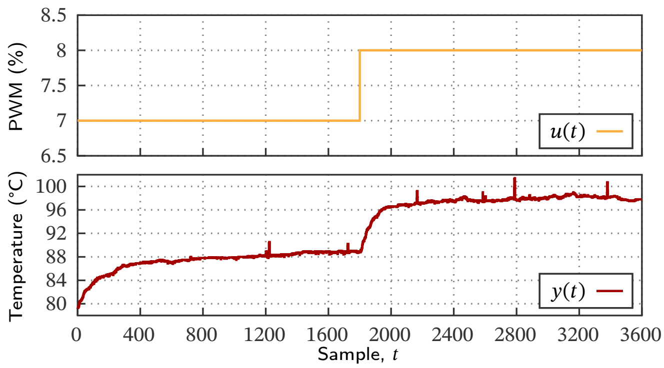


Figure 1. Open-loop data collected.

The controller's parameters are estimated using the OLS and DLS solutions, considering $L(q) = T_d(q)$, presented in Sections 3.2 and 3.3, respectively. The input-output signals are preprocessed to remove the offsets, and the trending was also removed from the output signal. The data used are either the entire data set or the relevant subsets selected by the criteria, presented in Section 4.

Figure 2 presents through the blue continuous line the first difference of the SSV, $\Delta\underline{\sigma}^2(\Phi_L(t))$, calculated using the data collected during the experiment. This figure also

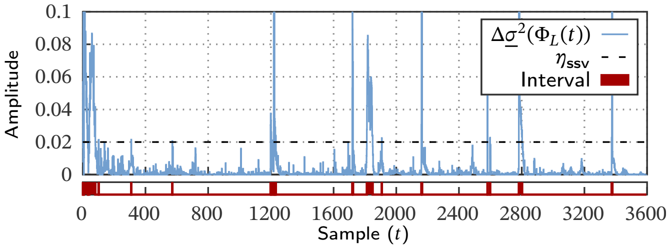


Figure 2. Evolution of $\Delta\underline{\sigma}^2(\Phi_L(t))$, threshold, and selected intervals.

presents the threshold chosen $\eta_{\text{ssv}}$ (black dashed line) and the relevant interval (red vertical lines). As mentioned before, one may use this plot to choose a reasonable value for $\eta_{\text{ssv}}$. Recall that low values of $\Delta\underline{\sigma}^2(\Phi_L(t))$ are produced by the rows of $\Phi_L$ that are strongly affected by noise. Observe in Figure 2 that it is possible to choose a threshold in order to avoid the small peaks. However, notice in this same figure that $\Delta\underline{\sigma}^2(\Phi_L(t))$ presents large values which do not come from the input signal exciting the process, but from the measurement errors in the output signal. Therefore, those rows of $\Phi_L$ are misclassified as informative by the criterion. Here, it was chosen $\eta_{\text{ssv}} = 0.02$ reducing $\Phi_L$ to about 164 relevant lines, that is, the parameters are estimated using 4.56% of the entire data set.

In a similar way, the CN was calculated using the data collected during the open-loop experiment. Figure 3 presents $\kappa(P(t))$ through the blue continuous line, along with the threshold $\eta_{\text{cn}} = 60$ (black dashed line) and the relevant intervals (yellow vertical lines). Observe in this figure that the CN not only increases with time, there are parts in which it actually decreases. That is the effect of the measurement noise. However, those samples are not classified as informative because the criterion is applied individually to each step change and uses the samples while the CN is smaller than the threshold. With the threshold chosen, 220 rows of $\Phi_L$ were classified as informative, meaning that the controller's parameters are estimate using 6.11% of the entire data set.

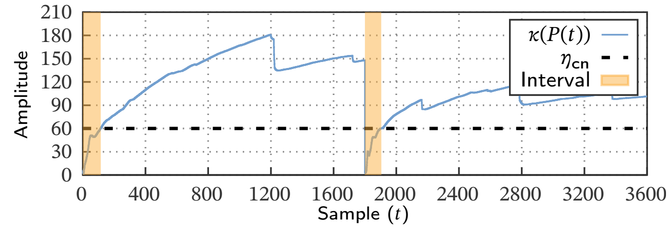


Figure 3. Evolution of $\kappa(P(t))$, threshold, and selected intervals.

Table 1 presents the parameter vector estimated in each case with the OLS solution. Observe that, the parameters obtained using the entire data set and the relevant intervals selected with the SSV criterion are quite similar. The reason relies on the fact that the SSV criterion misclassified as relevant the measurement error data. Table 2 presents the

Table 1. Controller's parameters estimated in each case with OLS.

| $\rho$ | All data | SSV | CN | Combined |
|---|---|---|---|---|
| $K_p$ | 0.1130 | 0.1176 | 0.5255 | 0.4532 |
| $K_i$ | 0.0330 | 0.0325 | 0.0296 | 0.0335 |

controller's parameters estimated with the DLS solution and either the entire data set or the informative subsets selected with each criteria.

Table 2. Controller's parameters estimated in each case with DLS.

| $\rho$ | All data | SSV | CN | Combined |
|---|---|---|---|---|
| $K_p$ | 2.0049 | 1.2972 | 0.9675 | 1.0278 |
| $K_i$ | 0.0326 | 0.0179 | 0.0265 | 0.0245 |

Each set of parameters presented in Tables 1 and 2 were used to configure the PID controller attached to the process, where the derivative term was set to zero. For each controller a new closed-loop experiment was performed. During those experiments the reference signal was a step from 80 °C to 90 °C. Figure 4 shows the collected output $y(t)$ obtained with each controller estimated with the OLS solution along with the desired output $y_d(t)$. Observe the oscillatory behaviour obtained with the controller estimated with the entire data set and with the SSV criterion. This is expected because the SSV criterion misclassifies as informative the variations in the output signal produced by the measurement errors. On the other hand, notice that the responses obtained with the informative subsets selected with the CN criterion and combining the criteria are much more closer to the desired one than the ones obtained using the entire data set.

Figure 5 presents the output signal collected in each experiment performed with the controllers estimated with the DLS solution. This same figure also shows the desired output $y_d(t)$. Observe that the responses obtained with the selected subsets are closer to the desired one than the

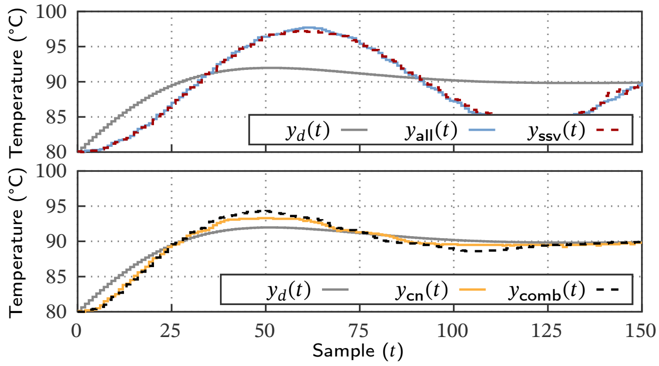


Figure 4. Closed-loop data collected with the controller's obtained with the OLS solution, and desired output.

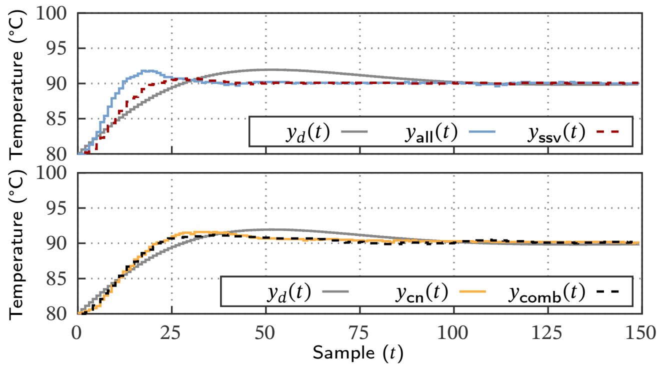


Figure 5. Closed-loop data collected with the controller's obtained with the DLS solution, and desired output.

one obtained using the entire data set. Observe also that those responses do not present oscillatory behaviour. That occurs because DLS considers that the regressor matrix is affected by noise, which suits well the VRFT problem when using open-loop data, as mentioned before.

Finally, the results are numerically compared using an estimate of the performance criterion cost function given by

$$\hat{J}^{\mathrm{mr}}(\hat{\rho}) = \frac{1}{N}\sum_{t=1}^{N}\left[y(t,\hat{\rho}) - y_d(t)\right]^2,$$

where $\hat{\rho}$ is the parameter vector estimated with each criterion and solution (OLS and DLS) and $N = 150$, which is the number of samples collected in the experiment. Table 3 presents the values of $\hat{J}^{\mathrm{mr}}(\hat{\rho})$ calculated for each case. Observe that the smallest values were found using the CN criterion and combining the criteria, regardless the solution. Those results agree with the closed-loop responses obtained in Figures 4 and 5.

Table 3. Cost function estimated using each criterion and solution.

| | $\hat{J}^{\mathrm{mr}}(\hat{\rho})$ | | | |
|---|---|---|---|---|
| Criterion | All data | SSV | CN | Combined |
| OLS | 0.3181 | 0.3045 | 0.0745 | 0.1054 |
| DLS | 0.1279 | 0.0832 | 0.0629 | 0.0618 |

## 6. CONCLUSIONS AND FUTURE WORK

The data-driven control design usually requires data collected from a specific experiment to adjust the controller's parameters. A convenient option to avoid performing this specific experiment is to use data collected during routine operation. However, applying those raw data along with a data-driven method may lead to inappropriate tuning, as illustrated in the experimental results. In the present work, we have applied and combined, two data selection criteria to select the relevant intervals for use with the VRFT method. Besides, we have proposed to use the data least squares solution to tune the controller's parameters with VRFT. From the results, we conclude that it is advantageous to select the informative data. In fact, considering the OLS solution, we have obtained closed-loop performances approximately 3 (combined criteria) and 4 (CN) times closer to the desired one compared to using the entire data set. On the other hand, with the DLS solution, we have obtained closed-loop performances approximately 2 times closer to the desired one (combined criteria and CN) than using the whole data set. However, the SSV criterion did not present the same improvement, because it misclassified as informative the output signal variations produced by the measurement errors. We also conclude that it is preferable to use the selection criteria along with the DLS solution, because, it resulted in closed-loop performance closer to the desired one than with the original solution, even when using the entire data set.

As future work we intend to investigate how to automatically adjust the thresholds used in the criteria. Moreover, we also intend to apply the data selection criteria and the DLS solution to other DD methods.

## REFERENCES

Arengas, D. and Kroll, A. (2017). A search method for selecting informative data in predominantly stationary historical records for multivariable system identification. In *Proc. of the 21st International Conference on System Theory, Control and Computing (ICSTCC)*, 100–105. Sinaia, Romania.

Bazanella, A.S., Campestrini, L., and Eckhard, D. (2011). *Data-driven controller design: the H2 approach*. Springer Science & Business Media, Netherlands.

Bittencourt, A.C., Isaksson, A.J., Peretzki, D., and Forsman, K. (2015). An algorithm for finding process identification intervals from normal operating data. *Processes*, 3, 357–383.

Campi, M.C., Lecchini, A., and Savaresi, S.M. (2002). Virtual reference feedback tuning: a direct method for the design of feedback controllers. *Automatica*, 38(8), 1337–1346.

Carrette, P., Bastin, G., Genin, Y., and Gevers, M. (1996). Discarding data may help in system identification. *IEEE Transactions on Signal Processing*, 44, 2300–2310.

DeGroat, R.D. and Dowling, E.M. (1993). The data least squares problem and channel equalization. *IEEE Transactions on Signal Processing*, 41(1), 407.

Garcia, C.S. and Bazanella, A.S. (2019). Extraction of informative subsets from routine operating data for use in data-driven control. In *Proc. of the 58th Conference on Decision and Control (CDC)*, 8392–8397. IEEE.

Garcia, C.S. and Bazanella, A.S. (2020). Selection of informative intervals in routine operating data for use in data-driven control design. In *Proc. of the Conference on Control Technology and Applications (CCTA)*, 225–230. IEEE.

Gevers, M., Bazanella, A.S., Bombois, X., and Miskovic, L. (2009). Identification and the Information Matrix: How to Get Just Sufficiently Rich? *IEEE Transactions on Automatic Control*, 54, 2828–2840.

Novus (2021). Controlador n1200. online. URL `https://www.novus.com.br/downloads/Arquivos/manual_n1200_v20x_j_pt.pdf`.

Shardt, Y.A. and Huang, B. (2013). Data quality assessment of routine operating data for process identification. *Computers & Chemical Engineering*, 55, 19–27.

Van Huffel, S. and Lemmerling, P. (2013). *Total least squares and errors-in-variables modeling: analysis, algorithms and applications*. Springer Science & Business Media, Netherlands.