IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 28 February 2025, accepted 20 March 2025, date of publication 24 March 2025, date of current version 2 April 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3554187

RESEARCH ARTICLE

Set Membership Adaptive Non Parametric Identification of Non-Linear Systems

JUAN ALEJANDRO CASTANO[1], **FERNANDO QUEVEDO**[2], **(Member, IEEE), AND FREDY RUIZ**[3], **(Senior Member, IEEE)**

[1]Department of Applied Mathematics, Science and Material Engineering, and Electronics Technologies, School of Experimental Sciences and Technology, Best research gropup Rey Juan Carlos University, 28933 Madrid, Spain
[2]Department of Computer Science and Engineering, Carlos III University of Madrid, 28911 Madrid, Spain
[3]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20233 Milan, Italy

Corresponding author: Juan Alejandro Castano (juan.castano@urjc.es)

This work was supported in part by Italian Ministry of University and Research through ''Learning-Based Model Predictive Control by Exploration and Exploitation in Uncertain Environments'' (PROGETTO DI RICERCA DI RILEVANTE INTERESSE NAZIONALE (PRIN), PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) 2022 Fund) funded by Ministerio de Ciencia, Innovación y Universidades. MCIN/AEI/10.13039/501100011033 under Grant P2022EXP2W and Grant PID2021-123657OB-C32; and in part by the ''European Regional Development Fund A way of making Europe (ERDF): A way of making Europe,'' through the 2024 Call Funding for Projects Led by Emerging Researchers within the Agreement between the Community of Madrid and Universidad Rey Juan Carlos for the Promotion of Research and Technology Transfer (2023–2026), Spain, under Grant M3579.

ABSTRACT Adaptation is a desirable feature when dealing with the identification of complex systems. However, this property can be difficult to achieve when non-convex model structures such as neural networks are employed to parametrize the unknown system. This work introduces a novel approach for dynamically adapting the data set that defines the model in non-parametric Set Membership identification methods. The proposed solution constructs a nonparametric, nonlinear model of a discrete-time dynamical system by exploring the data set, assuming the system follows a Nonlinear Auto-Regressive model with exogenous Inputs (NARX) structure. The identification data are assumed to be affected by unknown but bounded noise. Specifically, two strategies are proposed to adapt the identification data set while preserving system performance dynamically. The first strategy allows the data set to incorporate new data as novel modeling information becomes available, while redundant information can be eliminated when memory conditions are reached. The second strategy introduces new information sequentially; once an auxiliary memory vector in the data set reaches its desired cardinality, the method orderly replaces the oldest data with newer dynamics. These strategies enable the identified models to adapt in response to unmodeled behaviors arising from time-varying dynamics or limited initial data sets, minimizing the need for extensive experimentations and allowing to dynamically reconstruct the data set for developing data-driven models. The effectiveness of the proposed approaches is demonstrated through the experimental modeling of a nonlinear mechatronic system. Performance is benchmarked against neural network models and a static Set Membership identification strategy. Results indicate that the proposed dynamic data set generation approach improves the accuracy and robustness of the model when using non-informative experimental data sets as starting point for the estimation, improving the overall performance of the data-driven modeling task and facilitating the use of these modeling techniques in real environments.

INDEX TERMS Nonlinear identification, data-driven modeling, adaptive modeling, time-varying systems, set membership identification.

I. INTRODUCTION

Data-driven modeling has been gaining interest when representing complex behaviors of systems in several areas [5], [9], [11], [18], [19], [21], [22], [23]. This technique allows the user to better analyze their models by taking advantage of

The associate editor coordinating the review of this manuscript and approving it for publication was Mouquan Shen.

their intrinsic capabilities. For example, in [1], the authors can manually introduce disruption to the model to verify the impact of specific scenarios applied to a supply chain allowing the study of possible scenarios. A different example of the flexibility in data-driven modeling can be found in [2], where the time delay is identified in communications, minimizing undesirable effects. Another example is given by [14], where a physic-informed Neural Network (*NN*) is used to overcome a lack of information by using physical laws, therefore, the data-driven identification will compensate the missing information by using the physical laws and border conditions imposed by the trained *NN* with a forgetting factor. These mentioned cases are just a few of many that can be found in the state of the art.

Even though data-driven modeling is a hot topic, several issues emerge under different circumstances related to the generation of adequate data sets, required to estimate the models. As mentioned in [14], collecting informative data that meets the physical requirements might be very difficult or costly. In particular, *"the nonuniform and multiaxial stress data required for nonparametric identification remains nearly unmeasurable"*. To overcome this problem, the authors of that work, trained a *NN* with the physical laws that model the mechanical fields such as displacement, deformation gradient, stress, and strain energy density function subject to satisfactory respecting the physics *laws of momentum balance, compatibility, constitutive relations, and boundary conditions*. In this way, the imposed physics bounds compensate for the missing dynamic information in the identification data. In this way, the used data set for the identification process might run a proper model when poor data are introduced knowing that the output has to meet the physics law requirements.

A different problem related to the estimation data set is shown in [25]. In that paper, the authors analyze the impact of having outliers and extreme data in the set. In particular, they discuss how these data might produce sub-optimal solutions given the required constraints. Therefore, they proposed to use a smaller yet informative set to perform the optimization and verify the constraints with the fully informative data set, in this way, whenever the system is far from the boundaries, the minimization problem will converge faster and produce a less conservative solution given the nature of the smaller data set, while bounds are yet satisfied. If the output is not bounded, more identification data are used, obtaining a more conservative response. However, the case where the data set contains limited information is not given. A different approach to managing abnormal data is considered in [21], in that case, wind farms and solar plants performance analysis might be affected by failures, maintenance, energy curtailments, system degradation, etc. When analyzing the system's behavior, these data must be detected and removed from the sets. Therefore, that work proposed a method to identify such data and minimize their effect on the model. These data are assumed to be sporadic, so the model remains static.

Another strategy to overcome the identification problem appears in [15], where the authors proposed the *data-driven computing* paradigm. In this paradigm, the authors use data-driven techniques to map, given the input, the state that is the closest to the satisfaction of the boundaries and conservation laws, rather than identifying the closest state available in the identification which is closest to the data set.

In [27], the signal noise problem is considered when extracting parametric partial differential equations as model dynamics. Given the variability in the data due to noisy measurements, the authors counterbalance those behaviors by using physics-informed *NN*s that allow them to compensate for unseen behaviors based on physical principles and minimize the effect due to measurement noise. Compared to the work by [14], the first one used the physics-informed *NN* to identify noise while the second used the *NN* to compensate for the lack of information. However, the principle is similar in both cases.

Finally, as mentioned in [22], having a proper identification implies a cumbersome measurement operation that ideally requires high-resolution data with sufficiently high precision; however, as explained by the authors, obtaining such data is time consuming and not always possible. In addition, having incomplete data might lead to poor robustness in the identification and ill-conditioned or even rank-deficient model matrices. To overcome these problems, the authors proposed a hybrid adaptive identification problem that uses truncated singular values representation to minimize the ill-defined matrix problem and allow the model to evolve.

As exposed before, handling information in the available data sets is crucial when performing identification, this becomes more relevant when the system has time-varying dynamics, as mentioned in [10], [13], [16], [24], and [26]. As mentioned by [8], a common practice to model a slow time-varying system is to describe the system as a composition of time-invariant systems for a measured time. However, when data-driven techniques are used, the data acquisition phase needs to cover all possible variations through time, being time-consuming or not viable. To overcome such a problem, in [8], the authors used b-spline functions for double smoothing over different realizations and global time. In this way, they reduced the model order and noise effects.

Another approach is given by [20], where the authors divide the system model into time-variant parameters and time-invariant ones, they use the probability of new data to feed within the time-invariant parameters to explore the possibility of having new values for the variant ones using stochastic compartmental models. In [24] instead, the authors used the Frequency Transform Regressive Method, with a forgetting factor, to track the time-varying parametrization of an aircraft model.

Another possibility is given in [12] where online dynamic mode decomposition is presented, In that work, the authors initialize the model using a batch of measured data, afterward, the model is refined using an extended dynamic, based on new data, over the original estimation.

As has been presented, in data-driven techniques, data acquisition, and model variations are important factors limiting the applicability of such modeling techniques. On this behalf, we are introducing and analyzing the effect of having dynamically generated identification data sets when using the data-driven non-parametric Set Membership identification method [4], [6], [7], [17]. This method generates a non-parametric, non-linear mathematical model of a discrete-time dynamical system based on the data set exploration. Even though previous physical information is not considered as in [14] and [15], the system is assumed to be a Nonlinear Auto Regressive model with exogenous Inputs (NARX), described by a Lipschitz function with bounded Lipschitz constant, so that the gradient of the system output concerning the input regressor is limited, and this limit can be learned from the identification data set as a physical model limitation. In addition, the identification data is assumed to be affected by unknown but bounded (NBB) noise. In that sense, in this paper, we propose a dynamic construction of the identification data set, so that unmodeled behaviors, due to either, time-varying dynamics or limited data sets, are overcome. One proposed method incorporates a fixed-vector regressor with a forgetting factor, which systematically integrates new dynamic information into a fixed-size data set. Once the set reaches full capacity, the oldest dynamics are sequentially replaced by the newest ones. Additionally, a scenario where the fixed set serves as the complete identification vector is examined. The second method we introduce evaluates the impact of dynamically integrating new dynamics while retaining all previously used information relevant to the current system state. This approach reduces the size of the dynamic regressor while preserving the essential information introduced. We compare the developed identification methods with the original set membership approach given in [17] and with a static *NN*. The results demonstrate an appropriate learning rate for the studied cases with proper fitting values across all the studied scenarios.

The paper is structured as follows, Sec. II, briefly describes the set membership method; Sec. III describes the dynamics sets strategies and implications; results are given in Sec. IV and finally conclusions are presented in Sec. V.

II. SET MEMBERSHIP IDENTIFICATION

This section provides a brief description of the non-linear Set membership identification method presented in [17].

A nonlinear system with a NARX (Nonlinear AutoRegressive with eXogenous input) structure is described as

$$y(t) = f_o(\boldsymbol{\omega}(t)), \tag{1}$$

where

$$\boldsymbol{\omega}(t) = \begin{cases} [y(t-1), \ldots, y(t-n_{y_0}), \\ u(t), \ldots, u(t-n_{u_0}), \\ u_1(t), \ldots, u_1(t-n_{u_1}), \\ \vdots \\ u_j(t), \ldots, u_j(t-n_{u_j})], \end{cases} \tag{2}$$

$$\boldsymbol{\omega}(t) \in \mathcal{R}^n, n = n_{y_0} + n_{u_0} + \cdots + n_{u_j}. \tag{3}$$

However, a Single Input-Single Output (SISO) system can be assumed without loss of generality. Even though the system described by f_o is unknown, a set of measurements $\widetilde{y}, \widetilde{\boldsymbol{\omega}}$ of the output y and the corresponding input regressor $\boldsymbol{\omega}$ is used to find an estimate $\hat{f}$ of f_o. The measurements are considered to be affected by a UBB measurement error $e(k)$, bounded by a constant ϵ, then

$$\begin{aligned} \widetilde{y}(k) &= f_o(\boldsymbol{\omega}(k)) + e(k), \\ k &\in [1:N], \end{aligned} \tag{4}$$

with $||e(k)|| \leq \epsilon, \ \forall k$. In addition, f_o is considered to be a Lipschitz continuous function with bounded gradient, that is

$$f_o \in \mathcal{F} \doteq \left\{ \begin{matrix} f \in C^1(\boldsymbol{\omega}) : \left\| f' \right\| \leq \gamma \\ \forall \boldsymbol{\omega} \in \boldsymbol{\Omega} \end{matrix} \right\}, \tag{5}$$

where $f'(\boldsymbol{\omega})$ denotes the gradient of $f(\boldsymbol{\omega})$ and $\|x\|$ is the Euclidean norm. Therefore, f_o belongs to the set of functions $\mathcal{F}$ whose first derivative norm is bounded by a factor γ, corresponding to the Lipschitz constant. Given the a priori information about the model and the noise, and the set of measurements $\Phi_0 = [\widetilde{y}(k), \widetilde{\boldsymbol{\omega}}(k)]$, $k \in [1:N]$, the **FSS** (Feasible System Set) of the problem is

$$FSS \doteq \left\{ \begin{matrix} f \in \mathcal{F} : |\widetilde{y}(k) - f(\widetilde{\boldsymbol{\omega}}(k))| \leq \epsilon, \\ k \in [1:N] \end{matrix} \right\}. \tag{6}$$

Then, the **FSS** includes all the functions (models) in $\mathcal{F}$ consistent with the prior information and the available measurements. i.e. it represents the set of models whose maximum gradient with respect to the regressor is bounded by γ, and its prediction error is lower than ϵ, and therefore are contained in this set, where the value of γ is estimated from the identification data set. If the hypotheses previously exposed are valid, $FSS \neq \emptyset$ and $f_o \in FSS$. Additionally, it can be demonstrated that, for a new regressor w, the optimal estimate of $f_o(w)$ is given by

$$f_c(\boldsymbol{\omega}) \doteq \frac{f_u(\boldsymbol{\omega}) + f_l(\boldsymbol{\omega})}{2}, \tag{7}$$

where

$$cf_u(w) = min_{1 \leq k \leq N}(\widetilde{y}(k) + \epsilon + \gamma \left\| w - \widetilde{\boldsymbol{\omega}}(k) \right\|), \tag{8}$$

$$f_l(w) = max_{1 \leq k \leq N}(\widetilde{y}(k) - \epsilon - \gamma \left\| w - \widetilde{\boldsymbol{\omega}}(k) \right\|). \tag{9}$$

From theorems 2, 5 and 7 in [17], it follows that

- $f_u(\boldsymbol{\omega})$ and $f_l(\boldsymbol{\omega})$ are optimal bounds for $f_o(\boldsymbol{\omega})$;
- $f_u(\boldsymbol{\omega})$ and $f_l(\boldsymbol{\omega})$ are Lipschitz continuous on W;

- f_c is an optimal estimate of f_o for any $L_p(\boldsymbol{\omega})$ norm, with $p \in [1, \infty]$;

where the optimality criterion is

$$cf_{opt} = \arg\inf_{\hat{f}} \sup_{f \in FSS} \left\| f - \hat{f} \right\|_p . \quad (10)$$

In the current setting, a non-empty FSS is said to validate the consistency of the a priori assumptions with the available measurements. If the noise bound ϵ and the Lipschitz constant γ are not known a priori it is necessary to validate the assumptions based on the measurements $\widetilde{y}, \widetilde{w}$ using Algorithm 1, as presented in [17], so that we find the minimum values of ϵ and γ such that $FSS \neq \emptyset$.

Algorithm 1 Validate Lipschitz Constant

```
function Update γ (Φ,ϵ)
    Define ϵ as an error percentage of the output or a
      specific value.
    Initialize auxiliary Lipchitz constants.
    γ=0;
    N=Cardinality(Φ);
    for k = 1 to N do
        for i = 1 to N do
            Evaluate the distance between ω(k) and ω(i)
              i ≠ k
            if i ≠ k then
                D = ‖ω̃(k) − ω̃(i)‖
                find γ for the upper bound as
                γ_u = |(ỹ(k) − ỹ(i) + ϵ)/D|
                if γ_u > γ then
                    γ = γ_u;
                end if
            end if
        end for
    end for
    return γ
end function
```

Remark 1: If the available data set does not contain information about all the dynamics of the systems due to experimental limitations, time-varying parameters, or similar, the true function that describes the system's dynamics might have a larger Lipschitz constant γ, therefore, it does not belong to the FSS and a greater γ should be used, otherwise the system output has limited dynamics and the output is valid only for operational points consistent with the identification data set construction experiment.

Knowing the Lipschitz constant γ, the estimation of the system output for the new regressor $\widetilde{\boldsymbol{w}}$ is done using Algorithm 2, with known computational complexity [7].

Remark 2: The set membership model is a non-parametric, mathematical model that uses the a-priori hypothesis on γ and ϵ, and the identification data to derive function estimates for unseen regressor values.

Remark 3: The computational complexity in Algorithm 2 depends on the amount of data in the identification data set.

Algorithm 2 NLSM

```
  For a new input regressor x
function NLSM(x,Φ,γ,ϵ)
    f_u(x) = +∞
    f_l(x) = −∞
    for k = 1 to N do
        Evaluate the distance between x and ω(k) as:
        D = ‖x − ω(k)‖.
        Obtain the Upper Bound, UB on f_o(x)
        UB = y(k) + ϵ + γD.
        Obtain the lower bound, LB on f_o(x)
        LB = y(k) − ϵ − γD.
        if UB ≤ f_u(x) then
            f_u(x) = UB
            save the used data for the upper bound.
            Index_u = k
        end if
        if LU ≥ f_l(x) then
            f_l(x) = LU
            save the used data for the lower bound.
            Index_L = k
        end if
    end for
    Compute model output.
    f_c(x) = (f_u(x) + f_l(x))/2

        return f_c(x), Index_L, Index_u
end function
```

So, a small data set with as much information as possible is desirable.

III. SET MEMBERSHIP IDENTIFICATION BASED ON DYNAMIC DATA SET GENERATION

As previously mentioned, in many applications it is difficult, costly, demanding, and even impossible, to run experiments that provide full information on a given system. In addition, for time-variant systems, this premise becomes normative and difficult for the application of data-driven identification methodologies. To cope with this problem, in this paper, we introduce three different dynamic data set generation methods that allow using the set membership models in a dynamic setting.

A. GENERALITIES

As mentioned previously, given a FSS as in (6) that has a Lipschitz constant γ validated over experimental data using Algorithm 1 as in [17]. Then the function $f \in \mathcal{F}$ that contains the desired dynamics is consistent with the dynamic response provided by $\hat{f}$ for a certain regressor $w(t)$ obtained from the experimental data using Algorithm 2 $\iff$ the dynamics of f_o has a Lipschitz constant smaller than γ. Therefore, if the nature of the experiments or systems introduces faster dynamics $\mathcal{F}$ no longer contains f_o.

At any time instant $k > N$, for a new sample of the system input-output data, $\boldsymbol{\omega}(k)$, $y(k)$, let $\nabla y(k)$ be:

$$\nabla y(k) = \max_{i=1:N} \frac{|y(i) - y(k)| + \epsilon}{\|\boldsymbol{\omega}(i) - \boldsymbol{\omega}(k)\|}. \tag{11}$$

If $\nabla y(k) > \gamma$, it implies that the existing FSS does not hold information for such dynamics, therefore $\hat{f}(\boldsymbol{\omega}(k)) - f_o(\boldsymbol{\omega}(k)) > |\epsilon|$, and the model output does not longer represents the system dynamics. In such a case, it is necessary to reconstruct the FSS to validate the model parameters with the new identification data set and provide a new $\mathcal{F}$ with a larger Lipschitz constant $\gamma_1 \geq \nabla y(k)$.

For a time instant $K + Q$, consider a new input regressor $\widetilde{\boldsymbol{\omega}}_{(K+Q)}$, using 2 we can evaluate $\hat{f}_0(\widetilde{y}_{(k+Q)})$. However, for such regressor, it is obtained that $norm(f_0(\boldsymbol{\omega}(n)) - \hat{f}_0(\widetilde{y}) > \epsilon$, therefore $\nabla(k + Q) = \gamma_1 > \gamma$. This implies that $\hat{f}_0$ no longer contains the model information consistent with prior information ϵ. In addition, the function $\hat{f}_0$ no longer represents the model and it follows that $FSS(\epsilon, \gamma, \Phi) = \emptyset$.

Hypothesis 1: By adding the pair of measurements $\widetilde{\boldsymbol{y}}_{(k+Q)}, \widetilde{\boldsymbol{\omega}}_{(k+Q)}$ to the identification data Φ with updated size $N + 1$ and $\gamma_1 > \nabla(y(k + Q,), \boldsymbol{\omega}(\boldsymbol{k}\,\mathcal{C}\,\boldsymbol{Q}))$, it is possible to update the FSS such that the function

$$f_0 \in \mathcal{F}_1 \doteq \left\{ f \begin{matrix} \in C^1(\boldsymbol{\omega}) : \|f'\| \leq \gamma_1 \\ \forall \boldsymbol{\omega} \in \boldsymbol{\Omega} \end{matrix} \right\}, \tag{12}$$

$$FSS \doteq \left\{ \begin{matrix} f \in \mathcal{F}_1 : |\widetilde{y}(k) - f(\widetilde{\boldsymbol{\omega}}(k))| \leq \epsilon \\ k \in [1 : N + 1] \end{matrix} \right\}, \tag{13}$$

where

$$y(n) = f_0(\boldsymbol{\omega}(n)) + e(n) \quad | \nabla f_1(\boldsymbol{\omega}(n)) \leq \gamma_1, \tag{14}$$

and the Lipschitz constant γ_1 is updated as in Algorithm 3 over the extended data set $[\Phi, [\widetilde{\boldsymbol{y}}_{(k+Q)}, \widetilde{\boldsymbol{\omega}}_{(k+Q)}]]$, then the reconstructed FSS fully described f_0 and $\nabla\hat{f}_1 \leq \gamma_1 \forall \boldsymbol{\omega}$.

If 1 is true, it is possible to regenerate the existing FSS by adding new data to the existing data set. However, theoretically, the data set might grow infinitely, therefore the additional data should be limited to guarantee proper execution and constraint the execution time according to the real-time constraints if required.

In this work, three different dynamic data set augmentation methods are given.

- **Dynamic identification set with forgetting factor**: Updates the identification data set with new information till a certain length. Once the maximum length has been reached the dynamic section of the regressor is self-validated using the described algorithm in [17]. Sec. III-B1
- **Dynamic identification set with memory cleaning**: Updates the identification data set with new information till a certain length. Once the maximum length has been reached the dynamic section of the regressor is rewritten from its older component (queue behavior). Sec. III-B2
- **Regenerative identification set**: Updates the identification data set with new information in a queue-type behavior. Sec. III-B3

Algorithm 3 Updated Lipschitz Constant

```
function Update γ (ỹ,x,Φ,γ,ε)
    where Φ = [ŷ(i), ω(i)]|i ∈ [1 : N]
    γ1 = γ2 = 0
    for a new input regressor x
    [ŷ, index_l, index_u] = NLSM(x, Φ, γ, ε); Algorithm 2;
    measures the system output at the next sample.
    if ‖(ỹ − ŷ)‖ > ζ then (Estimation error greater than desired)
        if ‖x − ω(index_u)‖ ≠ 0 then (Recompute γ for the upper bound)
            γ1 = (ỹ − ŷ(index_u) − ε) / ‖x − ω(index_u)‖;
        end if
        if ‖x − ω(index_l)‖ ≠ 0 then (Recompute γ for the lower bound)
            γ2 = (ỹ − ŷ(index_l) + ε) / ‖x − ω(index_l)‖;
        end if
        γ = max([|γ|, |γ1|, |γ2|]);
    end if
    return γ
end function
```

The first two consider the system to be time-invariant but the identification data set provides limited information on the full system dynamics. The third strategy is considered for time-variant systems, so the data set will be eventually modified.

B. DYNAMIC DATA SET FOR DYNAMIC SYSTEMS

This method extends the existing data set with the corresponding regressor $\boldsymbol{\omega}(k)$ for the measured $\widetilde{y}(k)$ when $\left\|(\widetilde{y}(k) - \hat{y}(k))\right\| > \zeta$, where ζ stands for the admitted error during the identification under which there is not required to add additional data to the identification data set. The admitted error is a decision variable defined by the user while the regressor at sample k, $\boldsymbol{\omega}(k)$, and the measured output, $\widetilde{y}(k)$, are assumed measurable.

Given the measured output

$$\widetilde{y}(k + 1) = f_o(\boldsymbol{\omega}(k + 1)) + e(k + 1), \tag{15}$$

and the estimation

$$\hat{y}(k + 1) = f_c(\widetilde{\boldsymbol{\omega}}(k + 1)), \tag{16}$$

if

$$\bullet\ \nabla y(k + 1) = \gamma_1 > \gamma. \tag{17}$$

$$\bullet\ \left|\widetilde{y}(k + 1) - \hat{y}(k + 1)\right| > \zeta. \tag{18}$$

Therefore, due to the first condition in (17), the existing FSS with prior information $[\gamma, \epsilon]$ is falsified and does not contain f_o, while the second condition in (17)is a relaxation of this conditioning. Then, data $[\widetilde{y}(k + 1), \widetilde{\boldsymbol{\omega}}(k + 1)]$ is added to the identification data set and the Lipschitz constant γ updated in agreement with [17] and Algorithm 3, allowing

$f_c|_{k+1}$ to belong to

$$lFSS_1 \doteq \left\{ \begin{matrix} f \in \mathcal{F} : |\widetilde{y}(k+1) - f(\widetilde{\boldsymbol{\omega}}(k+1))| \leq \epsilon \\ \epsilon \in [1:N+1] \end{matrix} \right\}, \quad (19)$$

and

$$cf \in \mathcal{F} \doteq \left\{ f \in C^1(\boldsymbol{\omega}) : \|f'\| \leq \gamma_1, \forall \boldsymbol{\omega} \in \boldsymbol{\Omega} \right\}. \quad (20)$$

As previously said, and as can be seen from Algorithm 2, having a longer identification data set, implies more computational time to perform a single estimation, therefore, according to the application, it might be required to limit the maximum size that the data set might get. Once the maximum length is reached, creating new space or updating the updated identification data set is necessary. In that sense, two approaches are given

1) MEMORY CLEANING

The first method we introduce consists of concatenating an existing data set Φ_0 with new data representing unmodeled dynamics. The new data $[\widetilde{\boldsymbol{\omega}}(N+k), \widetilde{y}(N+k)]$ are added to the new Φ_1 regressor every time $\left\|(\widetilde{y}(N+k) - \hat{y}(N+k))\right\| > \zeta$. Therefore, $\Phi = [\Phi_0 \bigcup \Phi_1]$ where Φ_0 contains the initial identification data and remains static, while the Φ_1 contains all the additional data that are being introduced during system execution. To deal with the increasing complexity that affects the execution time, the maximum amount of data must be restricted. Let the maximum cardinality of Φ_1 be $Card_1$, if $|\Phi_1| > Card_1$, a redundant information cleaning process is performed as in Algorithm 4 over Φ_1. This allows the data set to self-validate the membership hypothesis for the last included dynamics. i.e. remove all data in Φ_1 that do not impose an active upper or lower bound over all data in the same set. For each data in the Φ_1 auxiliary data set, we apply Algorithm 2 and obtain the indexes $index_L$, $index_u$ which indicate the samples' positions in Phi_1 that generate the lower and upper bounds for that sample. , leading to

$$c\Phi = \left[\Phi_0, \Phi_1|_{CleanID_{Set}}\right], \quad (21)$$

such that

$$cf \in \mathcal{F} \doteq \left\{ \begin{matrix} f \in C^1(\boldsymbol{\omega}) : \|f'\| \leq \gamma_1 \\ \forall \boldsymbol{\omega} \in \boldsymbol{\omega} \end{matrix} \right\}. \quad (22)$$

Notice that the same algorithm might be used to reduce the original regressor size, increasing the computational analysis in the original NLSM method as well.

Drawback: If no data is cleaned from the $ID\_set(k)$ after the cleaning process and new data needs to be entered in the regressor, the method will require to enter repetitively to the cleaning algorithm and the computational time will be affected.

To avoid such a situation, the length of the desired vector can be extended dynamically when the cardinality of Φ_1, after cleaning, is bigger than a certain percentage of the $Card_1$. In this way, the regressor will grow to allow new information to enter the system but decrease to its original size once redundant information enters the system.

2) FIX SIZE VECTOR WITH FORGETTING FACTOR

This approach considers the dynamics of the system to be evolving in time, therefore, once new dynamics invalidate the learning factor ζ, that is

$$c \left\|(\hat{y}(k+1) - \widetilde{y}(k+1))\right\| > \zeta, \quad (23)$$

the couple $[\widetilde{y}(k+1), \omega(k+1)]$ is added to Φ_1 including the new dynamics into the auxiliary regressor Φ_1. Once $|\Phi_1| = Card_1$, the new data is written from the beginning of Φ_1. Thus, old dynamics information is overwritten with the most recent data.

As reflected in Algorithm 5, the method assumes that the last added data to the regressor will appropriately model future dynamics. In contrast, the oldest data, which had been replaced with new information, provides redundant information to f_c or contains information over obsolete dynamics that no longer represent the actual system dynamics.

$$cFSS_0 \doteq \left\{ \begin{matrix} f \in \mathcal{F} : |\widetilde{y}(k+1) - f(\widetilde{\boldsymbol{\omega}}(k))| \leq \epsilon \\ k \in [1:N] \end{matrix} \right\}, \quad (24)$$

and

$$f \in \mathcal{F} \doteq \left\{ \begin{matrix} f \in C^1(\boldsymbol{\omega}) : \|f'\| \leq \gamma_0 \\ \forall \boldsymbol{\omega} \in \Phi \end{matrix} \right\},$$

where

$$\Phi = [\Phi_0, \Phi_1]. \quad (25)$$

3) FULLY ADAPTIVE REGRESSOR

In the previous case, Φ_0 represents a dynamic inherent to the system and does not change in time. It can be said that this dynamics stands for an operation point around which

Algorithm 4 *CleaningID_set*

```
function Φ = CleanID_set(Φ,γ,ε,Card_1)
    N=|Φ|;
    Φ_1 = ∅;
    for i = 1 to N do
        (yout,index_u,index_l)=NLSM(ω(i),[Φ\Φ(i)],γ,ε):
        usedIndex(i,1)=index_u
        usedIndex(i,2)=index_l
    end for
    for i = 1 to N do
        if i ∈ usedIndex then
            Φ_1 = [Φ_1 ⋃ Φ(i)];
        end if
    end for
    M = |Φ_1|
    if M > Card_1 * μ then (where μ < 1, as a percentage
        of Card_1)
        NewSetSize = Max(Card_1, M*ρ); (where ρ > 1 is
            an inflation coefficient).
    end if
    return [Φ_1, NewSetSize]
end function
```

Algorithm 5 FixSize With forgettingFactor

function $\Phi_1 = Fix_{set}(\Phi_0, \Phi_1, \gamma, \epsilon, Card_1, \omega)$
  For a new input regressor ω
  $\Phi = [\Phi_0 \bigcup \phi_1]$
  N=$|\Phi|$;
  for $i = 1$ to N **do**
    ($yout$,$index_u$,$index_l$)=NLSM(ω,Φ,γ,ϵ);(Algorithm 2)
    measure the system output at the next sample.
    if $\|(\tilde{y} - \hat{y})\| > \zeta$ **then** (Estimation error greater than desired)
      if $|\Phi_1| = Card_1$ **then**
        index=0;
      end if
      $\Phi_1(index + 1) = [\omega, \tilde{y}]$;
    end if
  end for
  return $[\Phi_1]$
end function

the identification data were acquired. On the other hand, Φ_1 represents the time-variant dynamics when the system moves far from the initial operation point, which might lead to a modification of the dynamics.

Suppose the variation of the dynamics among operation points is large. In that case, it can be better to assume an empty initial identification data set $\Phi_0 = \emptyset$ and rely only on the adaptive part of the set Φ_1, such that $\Phi = [\Phi_0 \Phi_1] = \Phi_1$ where Phi_1 is initialized with cardinality $|\Phi_1| = N$ and $N \leq Card_1$.

Modifying the vector from the first to the last position will add the new dynamic information orderly and keep track of the older dynamic information. Therefore, by modifying the oldest data with the incoming new dynamics, the added information will keep the proper information around the actual dynamics.

Remark 4: In general, it is assumed that a bigger constant γ will better contain the system dynamics, however, if the overall dynamics radically change because the system has new inertia, friction, operation points, etc. A smaller γ value may be required. Therefore, running the Algorithm in [17] is advisable. Notice that this computational time will affect the sample's estimation time. Therefore, this process should run in parallel or as a low-priority task such that the estimation time holds.

Remark 5: When validating γ in runtime over the updated data set, for cases in Sections III-B1 and III-B2 the initial condition γ equal to the one that validates the original FSS can be used, this requires γ to be validated only for the extended data set, minimizing the computational time. Instead, for the case in Section III-B3, such validation should be performed over the updated full regressor vector using initial condition $\gamma = 0$ since the new dynamics might require a smaller Lipschitz constant γ.

IV. RESULTS VALIDATION

In this section, the proposed methods are evaluated on several data sets generated by a rotary flexible manipulator produced by *Quanser Inc.* A large experimental campaign has been executed, with experiments considering varying complexity inputs. Parametric variations are performed on the system to generate data sets with different dynamic characteristics.

FIGURE 1. Rotary flexible Joint by Quanser innovate. Educate.

A. PLANT DESCRIPTION

The plant is a mechatronic system formed by two main bodies. A base, driven by a DC motor, and an arm of variable length. The two subsystems are connected through a flexible joint formed by a couple of springs, as shown in Fig. 1. The arm of the manipulator is formed by two segments, rigidly assembled by screws. The total length of the arm can be varied between a minimum of 29.8 *cm* and a maximum of 40*cm*, with a consequent variation of its moment of inertia.

The dynamic state of the system can be described by the angular displacement and velocity of the motor shaft θ and the deflection angle and speed of the arm with respect to the base α, as indicated in Fig. 1. The main non-linearity of the system is the dependence of the force generated by the springs on the deflection angle of the arm α. For a fixed arm length and neglecting gearbox undesired phenomena (such as static friction, backlash, and elasticity), the system can be described by a fifth-order lumped-parameters model and further simplified as a fourth-order model if the electrical dynamics of the motor inductance are neglected.

To generate experimental data for the identification process, the system is operated in closed loop by a proportional position controller that regulates the position of the motor shaft θ. Different position references are applied, including pseudorandom binary Signals (PRBS), chirp signals (increasing frequency sinusoidal), and Random Gaussian Signals (RGS). Quadrature encoders measure the angular displacements with a resolution of 4096 counts per revolution, while the digital control platform directly logs the voltage applied to the DC motor.

B. DESCRIPTION OF EXPERIMENTS AND DATA SETS

12 experiments were conducted using 4 different inertia configurations from 28.8 *cm* to 40*cm* as depicted in Fig. 1. The experimental input signal varies the motor position reference as: pseudo-random binary sequence (PRBS) $\pm 2.5°$,

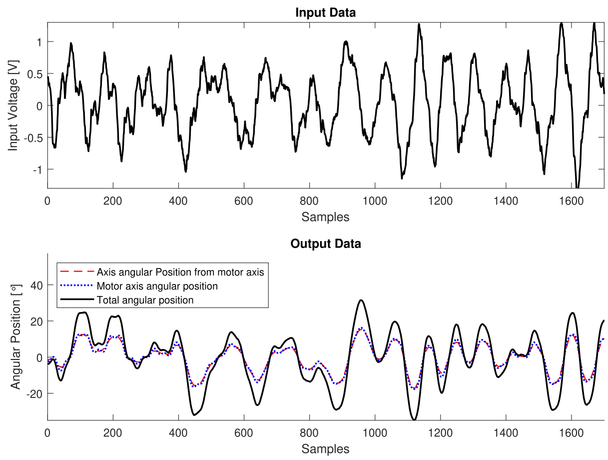

FIGURE 2. Input/Output system data for an RGS input and short arm.

TABLE 1. Data distribution for identification and validation for low inertia experiments.

	Identification	Validation
Chirp	1151	1251
PRBS	651	751
RGS	801	901
TOTAL	2603	2903
TotalValidationData		11612

Chirp $\pm 8°$, and random binary signal ± 16. All input signals were filtered at 0.2% of the bandwidth. The data were acquired at $2ms$, however, for the identification data generation a resampling step at $10ms$ was applied. In Fig. 2, an acquired signal example is shown, the output signal is the sum of the motor axis output and the arm angular position measured from the motor's axis.

After collecting the data and creating the input-output regressors, the experimental data set for each inertia case, consists of 2,402 samples for the chirp signal, 1,402 samples for the RRBS signal, and 1,702 samples for the RGS signal. Each experimental set was partitioned into identification and validation sets, as detailed in Table 1. The identification set for the studied scenarios concatenates the corresponding identification sets accordingly. The validation set is the concatenation of all validation sets to simulate the introduction of non-modeled dynamics in the considered scenarios.

different working scenarios are studied to compare the performance of the proposed methods. In particular, we evaluate our results against the standard *NLSM* and a *NN*. These models were selected based on two key factors. First, the NLSM method is the foundational model for implementing the proposed dynamic data set strategies. Second, NN is a well-established and widely used approach for nonlinear dynamic system identification as referred in [3], providing a robust baseline for comparison and facilitating a clearer interpretation of the given results.

For all the models the regressor is formed as:

$$\omega(t) = [V(t-i), \theta(t-j) + \alpha(t-j) | i \in [1, 4], j \in [1, 5]], \tag{26}$$

where $V(t)$ stands for the input voltage applied at time t and $\theta(t) + \alpha(t)$ is the absolute output shaft position of the axis, including the motor displacement, at time t.

The following five data-driven models are identified and compared:

- **NLSM**: Original NonLinear Set Membership model with initial Lipzchitz constant $\gamma = 0$ and $\epsilon = 10\%$.
- **NLSM$_{\mathbf{FRS}}$** from Algorithm 4. Dynamic set with forgetting factor in Section III-B1, extended regressor size: 100, $\zeta = 5\%$.
- **NLSM$_{\mathbf{CR}}$** from Algorithm 5. Dynamic set with fix vector in Section III-B2, extended regressor size: 100, $\zeta = 5\%$.
- **NLSM$_{\mathbf{NR}}$** from Algorithm 5 with $\Phi_0 = \emptyset$. Dynamic set for time-variant systems in Section III-B3, $\zeta = 5\%$.
- **NN**: Nonlinear Autoregressive Neural Network with External Input, with structure defined as input delays = 1, feedback delays = 1; hiddenLayerSize = 20; training data: 50% validation Radio: 25% and test radio: 25%; Optimization algorithm: Levenberg-Marquardt (LM); Loss function: Mean Squared Error.

C. INFORMATIVE EXPERIMENTAL DATA FOR LOW INERTIA CONDITION

In this case, the identification data set concatenates the three types of input signals, previously described, and only under the low inertia configuration, which means the link was extended to its minimum, at 29.8 cm, see Table 1 In this way, $N = 2603$ input-output samples form the training set and identification sets, while the estimation set has $M = 11612$. The complete identification data set is used for training the *NN*, however, for the *NLSM*-based estimators, a reduced data set with $N = 981$ samples is derived after applying the cleaning Algorithm 4.

The experimental set contains most of the system dynamics, and all the identified models obtained provide an adequate estimation of the validation data set with the fitting values as presented in Table 2. With this experiment, we can validate the selected regressor and NN structure that provides a good estimation of the dynamic system. Additionally, the Mean Absolute Percentage Errors (MAPE), Root Mean Square Errors (RMSE), and Maximum Errors (MAXE) are analize.

As seen in Table 2, the proposed methods achieve the best fitting performance, while the original set membership algorithm exhibits the lowest performance. However, due to the highly informative nature of the identification data set, all models perform satisfactorily. Notably, in terms of error performance, the NLSM algorithm demonstrates a maximum error that is 4% higher than that of the other methods. Nevertheless, the overall fitting performance differs by only

2%, with a slight variation in the mean and RMS error percentages (≈ 0.5%).

TABLE 2. Fitting and error Values for the informative experimental case, Motion Range for the validation test = 36.82°.

Method	$NLSM$	$NLSM_{FRS}$	$NLSM_{CR}$	$NLSM_{NR}$	NN
Fit %	89.11	91.13	91.13	91	90.06
MAPE%	1.01	0.96	0.96	1	1.07
RMSE%	2.75	2.24	2.24	2.27	2.38
MAXE%	11.56	7.24	7.24	6.72	7.47

As seen from Fig. 3, all but the NLSM model have errors below 5%. This implies that the three proposed estimators learned the unmodeled dynamics below an error in the given bound ζ. In this way, a performance similar to that of the NN one is obtained. In this case, the provided training and Identification data sets, are quite informative and allow all the methods to give a proper estimation.

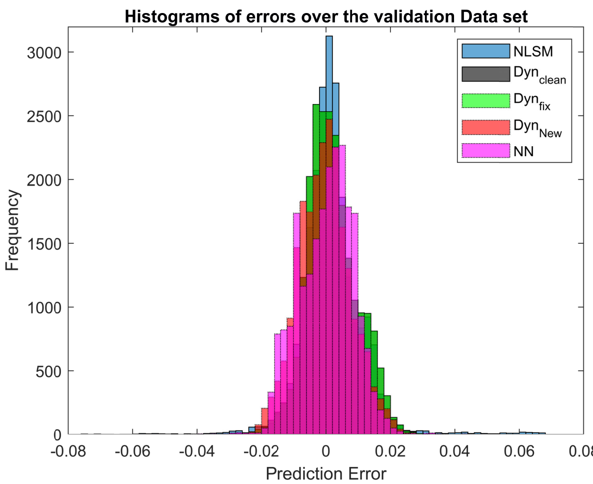

FIGURE 3. Histogram when using small inertia.

Analyzing the output performances in Fig. 4, it is seen that the $NLSM$, in the dashed line, does not reach the higher pick at 110.57 [s]. Given that adaptive regressor methods always depend on previous information, it is seen that once the $SMNL$ method reaches its saturation point, adaptive methods modify the estimation once the 5% error is reached. For this reason, a stair-like estimation behavior can be seen in the zoomed area in Fig. 4 at time 110.6 s.

D. LIMITED INFORMATION CASE USING LOW INERTIA CONDITION

In this case, we used as training data and identification sets, the measures from applying only a chirp signal to the system when using the lowest inertia configuration, i.e bar length 29.8 cm, see Table 1. $N = 1151$ input-output samples form the training set and identification sets, while the estimation data set contains $M = 11612$ samples including all experiments under the different working conditions. The complete identification data set is used for training the NN, however, for the $NLSM$-based estimators, a reduced data set with $N = 687$ samples is derived after applying the cleaning Algorithm 4. Therefore, the training set used during the identification is smaller than the previous one. A NN with the same structure previously used is employed.

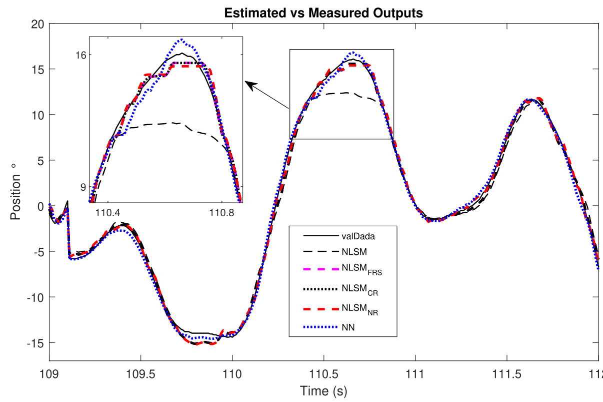

FIGURE 4. Output performances for the low inertia case.

TABLE 3. Fitting Values and errors for the low informative experimental case, Motion Range for the validation test = 36.82°.

Method	$NLSM$	$NLSM_{FRS}$	$NLSM_{CR}$	$NLSM_{NR}$	NN
Fit %	59.7	89.8	89.8	90.4	76.7
MAPE%	2.62	1.12	1.11	1.09	2.07
RMSE%	9.2	2.57	2.57	2.43	5.78
MAXE%	29.47	11.07	11.07	11.07	21.61

As shown in Fig. 5, the original $NLSM$ and the NN models trained with these data have many outliers during the validation. Yet, the $NLSM$ models when using the modified identification algorithms can learn the dynamics and maintain the identification errors below the expected $\zeta = 5\%$ as desired. In the zoomed area, we can visualize the histogram for the adaptive $NLSM$ models only. It can be seen that few outliers remain, however, most of the estimations stand inside the given error bound. It is important to mention that the NN results were given after training the NN multiple times, given that the optimization problem is non-convex and each time a different performance is obtained. The present one is the best one we obtained with the given data and structure. Note also that the adaptive models present a lower dispersion than the NN.

Looking at the system's output in Fig 6, it can be seen that the $NLSM$ using the original data set does not have enough information to identify the model when the output amplitude is big. Additionally, the NN has some miss predictions at time 83.2 [s]. In contrast, the methods using the dynamic regressor sets have a stair-like behavior in the rising and falling cycles. This performance is given on the first slope while the following ones present a smoother behavior as shown in the second zoomed area at time 88 [s] showing proper dynamic learning of the unmodeled dynamics. Therefore, once an unknown dynamic enters the system, the incorporated data can provide a good estimation for future system behaviors with similar conditions, as expected.

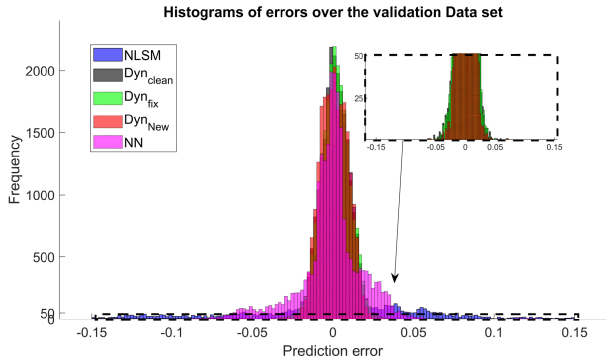

FIGURE 5. Histogram for Low Inertia non-informative case.

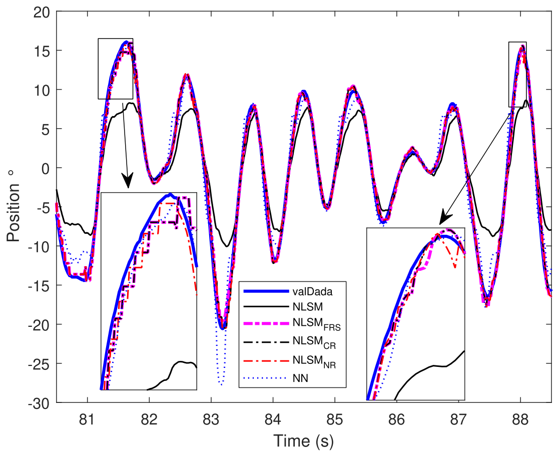

FIGURE 6. Estimated Output for the Low Inertia non-informative case.

Finally, as seen in Table 3, the fitting values when using dynamic regressor approximation perform better in low informative cases, showing proper learning of the system dynamics, as previously discussed. As presented, the three proposed methods for dynamic identification data sets provide better performances and can learn new information regarding unmodeled dynamics. Furthermore, even though the mean and RMS errors for the *NLSM* and NN are both smaller than 3%, which can be considered as small, the proposed methodologies reduced them to less than 1.12 % in the worst scenario and what more interesting, the maximum estimation error was reduced from 29.47 % for the *NLSM* and 21.61% for the *NN* methods to 11.07% for all dynamic regressor methodologies, this shows a proper and fast learning capacity. Notice that the maximum identification error exceeds the expected $\zeta = 5\%$, this might happen due to a fast dynamic introduction during the experiment transitioning.

E. PARAMETRIC PERFORMANCE COMPARISONS USING LOW INFORMATIVE AND LOW INERTIA CONDITION

Finally, without losing generality, we study the performance of one of the three proposed methodologies when varying the learning data conditions. In particular we evaluate the

TABLE 4. Fitting Values for different parameterizations of the set membership problem using strategy in Sec III-B2. Best *NN* fit value 76.7 [%], Motion Range for the validation test = 36.82°.

Vec_{lenght}	$\zeta\%$	Fit	$MAPE$ [%]	$RMSE$ [%]	$MAXE$ [%]
20	5	88.29	1.20	2.96	22.3
50	5	89.25	1.14	2.72	10.07
75	5	89.98	1.08	2.52	9.94
100	5	90.43	1.04	2.42	9.94
20	2	89.40	0.95	2.68	21.87
50	2	91.06	0.88	2.26	19.57
75	2	91.84	0.81	2.06	20.88
100	**2**	**91.95**	**0.79**	2.03	20.88
20	**8**	**87.33**	**1.28**	3.2	9.96
50	8	88.52	1.18	2.9	9.96
75	8	88.52	1.18	2.9	9.96
100	8	88.52	1.18	2.9	9.96

performance of the $NLSM_{CR}$. For this purpose, we used the data set in Sec. IV-D with $\epsilon = 5\%$ to validate the Lipchitz constant γ, and the total data was reduced using Algorithm 4. We vary the dynamic vector length Vec_{lenght} and the learning error factor ζ. Regarding the behavior w.r.t ζ, we analyze 3 cases, the first case considers that the bounded noise assumption is properly set and the system learns when the identification has an error greater than ϵ. The second case assumes a conservative bounded error ϵ, this implies that the resulting Lipchitz constant when validating the a-priori information in Algorithm 1 is smaller than required, due to a bigger bounded noise. Given the conservative assumptions, the algorithm learns when the identified error is smaller than the initial bound assumption. From a working perspective, the initial data set can determine the system with a maximum error ϵ. This data set is updated with fewer and more informative data allowing the convergence to a bounded error of 2%. In the third case, we modify the data set when the identification error surpasses the original assumption with a defined margin. This means learning only if a large dynamic error is observed.

As seen in Table. 4, the length of the learning vector has a greater impact on the performance by allowing more information to enter the data set. However, if the vector length is reduced, the update occurs constantly, making the method to forget past not-yet-old dynamics. When the learning error is bigger than the original hypothesis, the learning length loses weight as a parameter decision, since less information is required to model the unknown dynamics with bigger identification errors. In summary, the three methods are competitive, with small performance differences. As seen from the error point of view, the mean values are close to each other but the use of the dynamic regressor strategies reduced the maximum validation errors affecting the overall model performance.

If we compare the histograms in Fig. 7, which contains the worst and best cases presented in Table. 4 when using the *NLSM* with the dynamic regressor in Sec. III-B2, it is clear how the learning rate affects the mean errors and reduces the

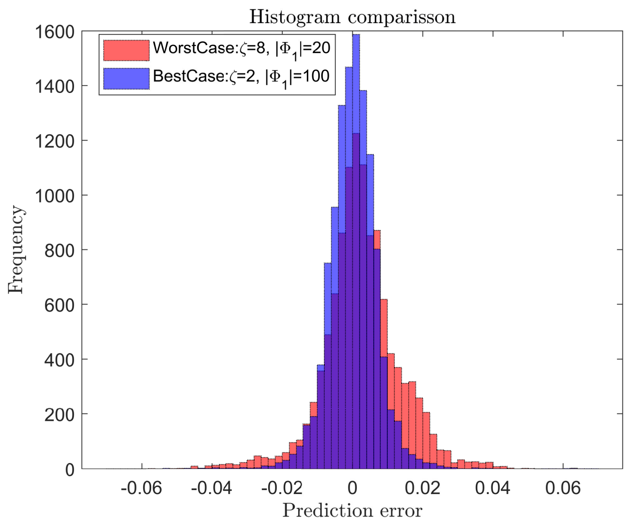

FIGURE 7. Histogram for best and worst cases.

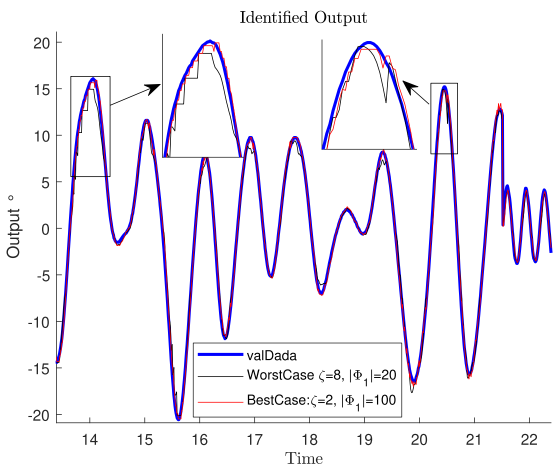

FIGURE 8. Best and worst cases output performances.

outliers based on the learning factor. However, it is important to note that the worst case, which has a $\zeta = 8\%$, can reduce the overall error inside the 5% band. New informative data with bigger errors introduces enough information to minimize the dispersion size.

In addition, Fig. 8 shows how the two studied cases adapt to the newly introduced dynamics at time 14 s with a stairs-like behavior once the error between real system and model outputs invalidates the parameter ζ. The best case scenario with $\zeta = 2\%$ and $|\Phi_1| = 100$ has a smaller step-like behavior w.r.t the worst case scenario with $\zeta = 8\%$ and $|\Phi_1| = 20$. Then, when similar dynamics is given at time 20.5 s, both models show a smooth response providing proper insight on the learned behaviors, as depicted in this area, both models respond smoothly but adaptations are required in both cases. However, the worst-case model decreases faster and requires bigger adjustments to align with the identification data. This behavior shows how a longer vector allows for a more informative regressor and its effect during a long-run

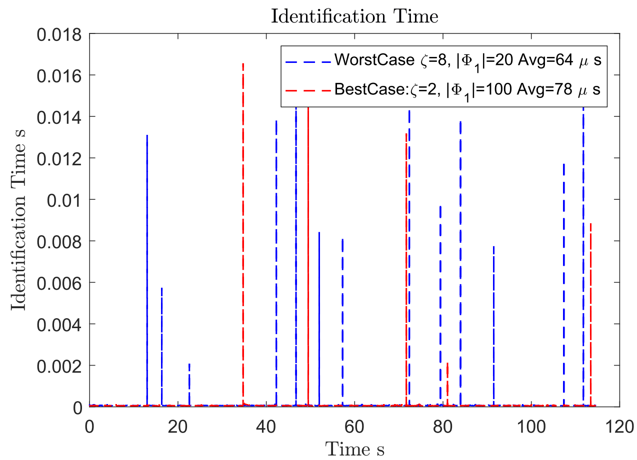

FIGURE 9. Identification time performance.

operation. Finally, as shown in Fig. 9, when using a smaller regenerative regressor, in the long run, it might request access far more times to the cleaning algorithm affecting the overall identification time which can be critical in some applications. This issue however, can be neglected by using a longer vector as shown in the figure, requiring a smaller updating frequency or in the best case, using parallel programming since the single data estimation for the analyzed cases is under 100 μs, while the estimation time when the cleaning data is called reaches up to 16 ms, which can be critical and restrictive in some cases.

According to the given results, when comparing the different methods the $NLSM_{FRS}$ allows the incorporation of new data into the model for estimation around new operational points, this method requires less computational resources than the $NLSM_{CR}$ proposed method, since the second one requires to execute a cleaning procedure to remove redundant information on the auxiliary regressor. Additionally, the $NLSM_{CR}$ allows for a dynamic size generation of the auxiliary data set which requires proper memory allocation and, in real-time applications, might compromise the system behavior. These two strategies allow the preservation of the information around specific working conditions extending the model usage around it. However, if the system is time-variant or no specific operational point is defined, the $NLSM_{NR}$ method, which is an extension of $NLSM_{NCR}$, replaces old dynamics with completely new information around any new operational point. This method is real-time save since operation quantification can be extracted as in $NLSM_{CR}$. From the operational point of view, according to the given results, the selection of the method is not critical since all of them have good adaptation to non-modeled dynamics in the data set. However, learning threshold ζ and auxiliary data set cardinality Φ play a major role in the performance.

V. CONCLUSION

This work presents a novel approach to dynamically generated data sets for data-driven set membership identification methods. The proposed methodologies are developed within a set membership non-linear non-parametric data-driven identification framework and were evaluated on a non-linear mechatronic system. Performance comparisons were presented considering neural networks as a test baseline. The proposed methodology addresses challenges like time-varying dynamics identification and limited identification data availability.

Three dynamic data set generation strategies are proposed. Memory cleaning, which allows a dynamic vector to grow till a certain cardinality and once reached, performs a cleaning of its elements to avoid redundant information. This strategy combines a priori experimental identification data set with a dynamic vector generation that adds new informative data to the data set when required. The second strategy has a forgetting factor that depends on the cardinality of an auxiliary memory vector, in that sense, once the vector is full, it is rewritten orderly forgetting the oldest data. Therefore, the method assumes older data to provide either redundant information or past dynamics. A particular use of such strategy is the *fully adaptive regressor* which stands when the auxiliary vector is initialized with the prior identification data set.

Results on a real non-linear system showed that using the proposed dynamic data set generation methods improved model accuracy and robustness by introducing meaningful information to the identification set under known working conditions, in particular noise bounds. The given results show improvements in RMS, mean, and maximum prediction errors with respect to the original *NLSM* and *NN* when low informative data sets are given. In case well informative data sets are available, the proposed strategies showed comparable performances to the based line methods here studied.

In addition, computational complexity is discussed and methods to limit the growing complexity due to the data set increasing length are discussed and analyzed, showing the existing relation between the auxiliary identification data set length and the final identification results. When using the *fully adaptive regressor* the proposed strategy proves to be effective with time-varying dynamics by continuously updating the identification data set to the new system behavior.

The proposed dynamic data set generation methodology enhances the accuracy of data-driven identification methods. It allows the dynamic to adapt the identified models to new information and handle time-varying dynamics. These characteristics make the proposed approaches useful for various applications in system modeling and identification. However, it is important to remark that under noisy conditions the learning parameter ζ can be invalidated by the noise variations, and then it results non useful, leading to the addition of redundant information to the data sets and subject to frequent invalidation of the model dynamics. In those cases, the additional data samples introduce components of the noise behavior in the identification data set. To prevent this, the experiments for acquiring the identification data set and the validation of hypothesis about ϵ must agree with the real working conditions. Additionally, if the system dynamics change continuously, and those changes are faster than the adaptation rate of the model, the proposed methods might not handle these adaptations properly and larger data sets might be required. Finally, these methods require higher computational resources w.r.t to the standard NLSM and other methods.

Future work includes further validation and their integration in adaptive control applications, potentially expanding their utility and impact in the field of data-driven modeling and dynamic control.

REFERENCES

[1] H. Aboutorab, O. K. Hussain, M. Saberi, F. K. Hussain, and D. Prior, "Adaptive identification of supply chain disruptions through reinforcement learning," *Expert Syst. Appl.*, vol. 248, Aug. 2024, Art. no. 123477.

[2] A. Bayrak and E. Tatlicioglu, "A novel adaptive time delay identification technique," *ISA Trans.*, vol. 139, pp. 156–166, Aug. 2023.

[3] F. Bonassi, M. Farina, J. Xie, and R. Scattolini, "On recurrent neural networks for learning-based control: Recent results and ideas for future developments," *J. Process Control*, vol. 114, pp. 92–104, Jun. 2022.

[4] M. Canale, L. Fagiano, and M. Milanese, "Set membership approximation theory for fast implementation of model predictive control laws," *Automatica*, vol. 45, no. 1, pp. 45–54, Jan. 2009.

[5] A. Carriero, M. Marcellino, and T. Tornese, "Blended identification in structural VARs," *J. Monetary Econ.*, vol. 146, Sep. 2024, Art. no. 103581.

[6] J. A. Castano and F. Ruiz, "Set membership identification of an excimer lamp for fast simulation," *Control Eng. Pract.*, vol. 21, no. 1, pp. 96–104, Jan. 2013.

[7] J. A. Castaño, F. Ruiz, and J. Régnier, "A fast approximation algorithm for set-membership system identification," *IFAC Proc. Volumes*, vol. 44, no. 1, pp. 4410–4415, Jan. 2011.

[8] P. Z. Csurcsia, J. Schoukens, I. Kollár, and J. Lataire, "Nonparametric time-domain identification of linear slowly time-variant systems using B-Splines," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 252–262, Jan. 2015.

[9] T. Drautzburg and J. H. Wright, "Refining set-identification in VARs through independence," *J. Econometrics*, vol. 235, no. 2, pp. 1827–1847, Aug. 2023.

[10] K. Dziedziech, P. Czop, W. J. Staszewski, and T. Uhl, "Combined non-parametric and parametric approach for identification of time-variant systems," *Mech. Syst. Signal Process.*, vol. 103, pp. 295–311, Mar. 2018.

[11] Z. Gao, X. Lin, and Y. Zheng, "System identification with measurement noise compensation based on polynomial modulating function for fractional-order systems with a known time-delay," *ISA Trans.*, vol. 79, pp. 62–72, Aug. 2018.

[12] M. A. Gultekin and A. Bazzi, "Real-time data-driven system identification of motor drive systems using online DMDc," in *Proc. IEEE Energy Convers. Congr. Expo. (ECCE)*, Oct. 2022, pp. 1–7.

[13] S. Huang, N. Lin, Z. Wang, Z. Zhang, S. Wen, Y. Zhao, and Q. Li, "A novel data-driven method for online parameter identification of an electro-chemical model based on cuckoo search and particle swarm optimization algorithm," *J. Power Sources*, vol. 601, May 2024, Art. no. 234261.

[14] I. Jeong, M. Cho, H. Chung, and D.-N. Kim, "Data-driven nonparametric identification of material behavior based on physics-informed neural network with full-field data," *Comput. Methods Appl. Mech. Eng.*, vol. 418, Jan. 2024, Art. no. 116569.

[15] T. Kirchdoerfer and M. Ortiz, "Data-driven computational mechanics," *Comput. Methods Appl. Mech. Eng.*, vol. 304, pp. 81–101, Jun. 2016.

[16] I. A. Kougioumtzoglou and P. D. Spanos, "An identification approach for linear and nonlinear time-variant structural systems via harmonic wavelets," *Mech. Syst. Signal Process.*, vol. 37, nos. 1–2, pp. 338–352, May 2013.

[17] M. Milanese and C. Novara, "Set membership identification of nonlinear systems," *Automatica*, vol. 40, no. 6, pp. 957–975, Jun. 2004.

[18] F. Quevedo, J. Muñoz, J. A. Castano Pena, and C. A. Monje, "3D model identification of a soft robotic neck," *Mathematics*, vol. 9, no. 14, p. 1652, Jul. 2021.
[19] F. Quevedo, J. M. Yañez-Barnuevo, J. A. Castano, C. A. Monje, and C. Balaguer, "Model identification of a soft robotic neck," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2020, pp. 8640–8645.
[20] B. Robinson, P. Bisaillon, J. D. Edwards, T. Kendzerska, M. Khalil, D. Poirel, and A. Sarkar, "A Bayesian model calibration framework for stochastic compartmental models with both time-varying and time-invariant parameters," *Infectious Disease Model.*, vol. 9, no. 4, pp. 1224–1249, Dec. 2024.
[21] H. Wang, N. Zhang, E. Du, J. Yan, S. Han, N. Li, H. Li, and Y. Liu, "An adaptive identification method of abnormal data in wind and solar power stations," *Renew. Energy*, vol. 208, pp. 76–93, May 2023.
[22] L. Wang, M. Li, G. Yu, W. Li, and X. Kong, "Automated measurement and hybrid adaptive identification method for kinematic calibration of hybrid machine tools," *Measurement*, vol. 222, Nov. 2023, Art. no. 113638.
[23] D. Wilson, "Data-driven model identification using forcing-induced limit cycles," *Phys. D: Nonlinear Phenomena*, vol. 459, Mar. 2024, Art. no. 134013.
[24] L. Xingju, G. Hongwu, and Z. Zhiqiang, "Online identification of time-variant parameters using the frequency transform regressive method," in *Proc. 34th Chin. Control Conf. (CCC)*, Jul. 2015, pp. 2099–2104.
[25] S.-B. Yang, S. Kammammettu, and Z. Li, "Data-driven distributionally robust chance-constrained optimization with large data set and outliers: Sequential sample removal algorithm for solution improvement," *Comput. Chem. Eng.*, vol. 179, Nov. 2023, Art. no. 108407.
[26] Y. Zhao, A. Fatehi, and B. Huang, "A data-driven hybrid ARX and Markov chain modeling approach to process identification with time-varying time delays," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 4226–4236, May 2017.
[27] H. Zhou, H. Li, and Y. Zhao, "Identification of partial differential equations from noisy data with integrated knowledge discovery and embedding using evolutionary neural networks," *Theor. Appl. Mech. Lett.*, vol. 14, no. 2, Mar. 2024, Art. no. 100511.

JUAN ALEJANDRO CASTANO received the B.S. and M.S. degrees in electronic engineering from Xavierian Pontifical University, Bogotá, Colombia, in 2011, and the Ph.D. degree in robotics, cognition and interaction technologies from Genoa University, Genoa, Italy, in 2016. Currently, he is an Assistant Professor with King Juan Carlos University, Madrid, Spain. His research interests include control techniques, system identification based on data drive algorithms, humanoid locomotion, and neurorehabilitation strategies with exoskeletons.

FERNANDO QUEVEDO (Member, IEEE) was born in Palma, Spain, in 1997. He received the B.S. degree in bioengineering from the Universidad de Málaga, in 2019. He is currently pursuing the master's degree in robotics and automatization with the University Carlos III of Madrid. He is collaborating with the Humasoft project researching models for soft robotics, 4-D path planners, and network control schemes for multi-UAV systems.

FREDY RUIZ (Senior Member, IEEE) was born in Facatativa, Colombia. He received the bachelor's and M.Sc. degrees in electronics engineering from Xavierian Pontifical University, Colombia, in 2002 and 2006, respectively, and the Ph.D. degree in computer and control engineering from the Politecnico di Torino, Italy, in 2009.

He was an Assistant Professor (2010–2014) and an Associate Professor (2015–2019) with Xavierian Pontifical University, where he was also the Head of the Electronics Engineering Department, from 2014 to 2016. He was a Fulbright Visiting Scholar with the University of California at Berkeley, in 2013; and a Visiting Professor with the Politecnico di Torino, in 2018. He is currently an Associate Professor with Politecnico di Milano, Italy. His research activity focuses on control and optimization, in particular the use and of data-driven techniques in optimal estimation and controller design, with applications in smart-grids, power electronics, robotics, and biotechnology.