Data-driven Nonlinear System Identification Under Bounded Noise Using Kernel Learning-based Ellipsoidal Set-membership

Hasna El Maizi[1,2], Mathieu Pouliquen[1], Rachid Fateh[2,3], Miloud Frikel[4], and Said Safi[2]

[1]*University of Caen Normandy, Caen, France,*
[2]*Sultan Moulay Slimane University, Beni Mellal, Morocco,*
[3]*Ibn Zohr University, Agadir, Morocco,*
[4]*École Nationale Supérieure d'Ingénieurs de Caen, Caen, France*

https://doi.org/10.26636/jtit.2026.4.2735

Submitted: 21 Jul. 2026, Accepted: 7 Sep. 2026

Abstract — Nonlinear system identification under bounded interferences remains a challenging problem when the noise bound is unavailable a priori. This paper presents a recursive kernel-based set-membership identification algorithm that combines the ellipsoidal outer-bounding (EOB) set-membership method with kernel learning to address this issue. Instead of identifying the nonlinear system directly in the input space, the proposed method exploits a reproducing kernel Hilbert space (RKHS), where the nonlinear estimation problem is represented as a linear regression model. Furthermore, a recursive data-driven procedure is introduced to estimate the noise bound simultaneously with the model parameters. Theoretical convergence properties of the proposed algorithm are determined. The performance of the proposed algorithm is validated through numerical simulation.

Keywords — ellipsoidal outer-bounding, kernel methods, nonlinear system identification, online noise bound estimation, set-membership identification

1. Introduction

The identification of nonlinear dynamical systems is a fundamental problem in systems and control, with applications ranging from industrial process and robotics to communication systems and biomedical engineering. The objective is to construct mathematical models from measured input-output data that precisely reproduce the system dynamics. Although linear identification methods are well established, they are generally inadequate for modeling systems with nonlinear dynamics.

To overcome these limitations, over the past decades numerous nonlinear system identification methods have been developed. Representative examples include block-oriented structures such as Hammerstein [1], [2] and Wiener [3], [4] models, benchmark nonlinear systems [5], as well as neural networks, radial basis function networks, and kernel-based learning techniques. Among these approaches, kernel methods have gained significant attention due to their ability to transform nonlinear estimation problems into linear ones in a reproducing kernel Hilbert space (RKHS).

The main idea of kernel learning is to map the input data into RKHS [6], [7], where nonlinear relationships can be represented through linear models without explicitly computing the feature mapping [8]. This property has led to the development of numerous kernel adaptive algorithms, including kernel recursive least squares (KRLS) [9], kernel affine projection algorithms (KAPA) [10], and kernel conjugate gradient (KCG) algorithms [11].

Most kernel adaptive algorithms are derived from stochastic optimization principles and assume that the statistical properties of the measurement noise are known [12], [13]. In practice, however, such information is rarely available, particularly in industrial environments where disturbances are often deterministic, time-varying, or poorly characterized.

An alternative is a set-membership identification (SMI) method, which relies on the assumption that the disturbances are unknown but bounded rather than statistically characterized [14], [15]. Under this assumption, the objective is to determine the set of all parameter vectors that are consistent with the available measurements and the bound of the noise. The feasible parameter set is commonly represented by polytopes or intervals. Although these descriptions provide guaranteed parameter uncertainty bounds, their computational complexity increases rapidly with the number of measurements and model parameters.

To reduce this computational overhead, an ellipsoidal approximations have been proposed as an efficient alternative. In particular, the ellipsoidal outer-bounding (EOB) algorithm recursively encloses the feasible parameter set within an ellipsoid, providing a compact uncertainty description with low computational complexity. Therefore, EOB algorithms have become an attractive solution for bounded-error identifica-

This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) License.
For more information, see https://creativecommons.org/licenses/by/4.0/

tion problems. However, existing EOB algorithms have been developed primarily for linear systems [16]–[19].

Recently, some ellipsoidal set-membership algorithms have been extended to nonlinear systems [6], [20]. These methods construct ellipsoidal regions to represent the uncertainty associated with the estimated parameter set. However, they generally assume that the noise bound is known a priori, an assumption that is often unavailable in real applications. Moreover, the size of the ellipsoid defining the uncertainty region is not explicitly addressed in these methods.

A comparison with representative existing methods, including KEOB [20], SM-KRLS [6], and the noise-bound estimation method reported in [21], is summarized in Tab. 1. The comparison highlights the main methodological differences between these methods and the proposed algorithm.

A remaining challenge in ellipsoidal set-membership identification algorithms for nonlinear systems is the lack of knowledge of the noise bound. In many cases, this quantity is unavailable and must be estimated directly from the observed data. Although the noise-bound estimation method in [21] addresses the estimation of the disturbance bound in the case of overestimation for linear systems, the effects of both overestimation and underestimation of the noise bound need to be considered and extended to nonlinear systems.

1.1. Main Contributions

This paper proposes a recursive kernel-based ellipsoidal outer-bounding algorithm for nonlinear system identification under unknown-but-bounded disturbances. The proposed approach integrates kernel learning with set-membership identification to provide accurate parameter estimation.

The main contributions of this paper can be summarized as follows:

- A kernel-based extension of the ellipsoidal outer-bounding algorithm is proposed for nonlinear system identification, where the nonlinear identification problem is transformed into a linear regression problem in the RKHS feature space.
- An ellipsoidal uncertainty region is recursively developed, and its size is minimized to provide a more precise representation of the uncertainty associated with the estimated parameters.
- Since the noise bound is often unknown in practical applications, a recursive data-driven procedure is introduced to estimate the noise bound jointly with the model parameters, eliminating the need for prior knowledge of the noise bound.
- The convergence properties of the proposed identification algorithm are established theoretically.
- The effectiveness of the proposed algorithm is demonstrated through numerical simulations on nonlinear benchmark systems, with comparisons against existing ellipsoidal algorithms for nonlinear identification.

The remainder of this paper is organized as follows. Section 2 formulates the nonlinear identification problem. Section 3 presents the proposed algorithm along with the recursive noise-bound estimation procedure. Section 4 provides numerical results illustrating the performance of the proposed method. Finally, Section 5 concludes the paper.

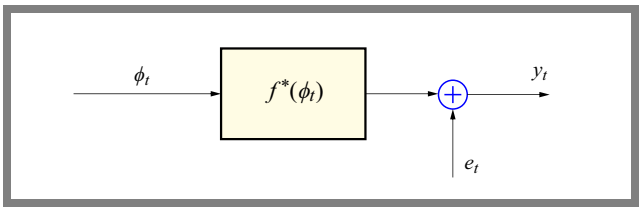


Fig. 1. System structure diagram.

2. Problem Formulation and Notation

Consider a nonlinear system represented by the following input-output equation (see Fig. 1):

$$y_t = f^*(\phi_t) + e_t \,, \tag{1}$$

where $\{y_t\}$ denotes the system output sequence and $\phi_t \in \mathbb{R}^m$ is the regression vector constructed from the current and past input-output samples, defined as

$$\phi_t = \begin{pmatrix} u_{t-1} \\ \vdots \\ u_{t-p} \\ y_{t-1} \\ \vdots \\ y_{t-q} \end{pmatrix} , \tag{2}$$

with $m = p + q$, where $\{u_t\}$ represents the system input sequence.

The function f^* is an unknown nonlinear mapping that characterizes the system dynamics, whereas $\{e_t\}$ denotes an unknown noise sequence.

To enhance the interpretability of the model, the function f^* is expressed as the sum of a linear component and a nonlinear component according to

$$f^*(\phi_t) = \phi_t^\top \theta^* + g^*(\phi_t) \,, \tag{3}$$

where $\theta^* \in \mathbb{R}^m$ is an unknown parameter vector associated with the linear part and $g^*(\cdot)$ represents an unknown nonlinear function associated with the nonlinear part.

In this paper, we assume the noise sequence $\{e_t\}$ is bounded, i.e., there exists a constant δ_e such that

$$|e_t| \leqslant \delta_e, \qquad \forall\, t \in \{1, \ldots, N\} \,, \tag{4}$$

where N denotes the number of available data samples.

Based on the parametrization and definitions introduced above, the proposed identification algorithms and their analysis are presented in the following section.

Tab. 1. Comparison of the proposed ellipsoidal identification algorithm with representative existing methods.

Feature	KEOB [20]	SM-KRLS [6]	Noise-bound estimation [21]	Proposed algorithm
System class	Nonlinear	Nonlinear	Linear	Nonlinear
Ellipsoidal uncertainty-region minimization	×	×	✓	✓
Noise-bound estimation	×	×	✓	✓
Convergence of parameter estimates	✓	✓	×	✓

3. Kernel-based Set-membership Identification Algorithm

3.1. Kernel Methods

Kernel-based learning provides an effective framework for modeling nonlinear systems by implicitly mapping the original input space into RKHS, where nonlinear relationships can be represented through linear operations. This transformation is achieved through a feature map

$$\psi : \mathcal{U} \to \mathcal{H} \,, \tag{5}$$

where $\mathcal{H}$ denotes a Hilbert space of functions and $\psi(\cdot)$ maps the input data into a possibly infinite-dimensional feature space.

In practice, the feature map is not explicitly computed. Instead, the inner products between feature vectors in $\mathcal{H}$ are evaluated through a kernel function

$$\kappa : \mathcal{U} \times \mathcal{U} \to \mathbb{R} \,,$$

satisfying

$$\kappa(u_1, u_2) = \big\langle \psi(u_1),\, \psi(u_2) \big\rangle_{\mathcal{H}} \,, \qquad u_1, u_2 \in \mathcal{U} \,. \tag{6}$$

This property enables the implicit manipulation of nonlinear mappings without explicitly constructing the feature vectors. Consequently, kernel methods extend linear estimation techniques to nonlinear problems while avoiding the computational burden associated with high-dimensional feature spaces.

Several kernel functions can be employed in practice, including the linear kernel $\kappa(u_1, u_2) = \langle u_1, u_2 \rangle$, the polynomial kernel $\kappa(u_1, u_2) = (\alpha + \langle u_1, u_2 \rangle)^q$, where $\alpha \geqslant 0$ and $q \in \mathbb{N}^*$, the sigmoid kernel $\kappa(u_1, u_2) = \tanh(\alpha \langle u_1, u_2 \rangle + \beta)$, the Laplacian kernel $\kappa(u_1, u_2) = \exp(-\|u_1 - u_2\|/\sigma)$, where $\sigma > 0$, and the Gaussian radial basis function kernel $\kappa(u_1, u_2) = \exp\big(-\|u_1 - u_2\|^2 / 2\beta^2\big)$.

The Gaussian kernel is the most widely adopted owing to its universal approximation capability and its effectiveness in capturing highly nonlinear relationships.

A reproducing kernel is formally defined as follows.
Definition 1. A function

$$\kappa : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$$

is called a reproducing kernel associated with a Hilbert space $\mathcal{H}$ of functions defined on $\mathcal{U}$ if it satisfies the following properties:

- For every $u \in \mathcal{U}$,

$$\kappa(u, \cdot) \in \mathcal{H} \,.$$

- For every $f \in \mathcal{H}$ and every $u \in \mathcal{U}$, the reproducing property holds:

$$f(u) = \big\langle f, \kappa(u, \cdot) \big\rangle_{\mathcal{H}} \,. \tag{7}$$

The second condition is referred to as the reproducing property, as it allows the evaluation of any function $f \in \mathcal{H}$ at a point u through an inner product in the Hilbert space.

The existence of a unique RKHS associated with a positive definite kernel is established by the following fundamental Theorem.
Theorem 1 (Moore–Aronszajn [8]). For every positive definite kernel κ defined on $\mathcal{U}$, there exists a unique reproducing kernel Hilbert space $\mathcal{H}$ for which κ is the reproducing kernel.

The Moore–Aronszajn theorem establishes the fundamental correspondence between positive definite kernels and RKHSs. Therefore, the selection of a kernel implicitly determines the functional space in which the unknown nonlinear function is represented.

A complementary characterization of positive definite kernels is provided by Mercer's theorem.
Theorem 2 (Mercer [8]). Let κ be a continuous symmetric kernel defined on a compact set $\mathcal{U}$. Then, κ is positive definite if and only if it admits the spectral decomposition

$$\kappa(u_1, u_2) = \sum_{m=1}^{\infty} \lambda_m \varphi_m(u_1)\, \varphi_m(u_2) \,, \tag{8}$$

where $\lambda_m \geqslant 0$ and $\{\varphi_m\}$ forms an orthonormal basis.

Mercer's theorem provides the theoretical foundation for the existence of an implicit feature representation and justifies the kernel formulation in Eq. (6). This result constitutes the basis of the kernel trick, which replaces inner products in the feature space by kernel evaluations, thereby avoiding the explicit computation of possibly infinite-dimensional feature vectors.

In the considered nonlinear system identification problem, directly estimating the unknown nonlinear component h^* in the original input space is generally difficult due to its nonlinear nature and potentially infinite-dimensional representation. To overcome this difficulty, kernel-based learning methods approximate g^* in the RKHS $\mathcal{H}$ associated with the kernel $\kappa(\cdot, \cdot)$.

According to the reproducing property, the nonlinear component $g \in \mathcal{H}$ satisfies

$$g(\phi_t) = \left\langle g, \kappa(\phi_t, \cdot) \right\rangle_{\mathcal{H}} . \tag{9}$$

For practical implementation, the nonlinear function is approximated using a finite kernel expansion constructed from a dictionary of L selected input samples $\{\phi_i\}_{i=1}^{L}$, leading to

$$g(\cdot) = \sum_{i=1}^{L} \theta_i^* \, \kappa(\phi_i, \cdot) \, , \tag{10}$$

where $\theta^* \in \mathbb{R}^L$ is the vector of kernel coefficients associated with the nonlinear component, defined as

$$\theta^* = \begin{pmatrix} \theta_1^* \\ \vdots \\ \theta_L^* \end{pmatrix} . \tag{11}$$

By defining the kernel-based regressor

$$\Psi_t = \begin{pmatrix} \kappa(\phi_1, \phi_t) \\ \vdots \\ \kappa(\phi_L, \phi_t) \end{pmatrix} \in \mathbb{R}^L, \tag{12}$$

and combining the linear decomposition in Eq. (3) with the finite kernel expansion in Eq. (10), the nonlinear system can be expressed as

$$y_t = \phi_t^\top \vartheta^* + \Psi_t^\top \theta^* + e_t \, . \tag{13}$$

Although the original system is nonlinear with respect to the input variables, the obtained model is linear with respect to the unknown parameter vectors ϑ^* and θ^*. This property is fundamental for applying linear estimation and set-membership identification methods to nonlinear systems through kernel-based representations.

The resulting nonlinear system model can then be rewritten in a form that is linear with respect to the unknown parameters:

$$y_t = \Phi_t^\top \Theta^* + e_t \, , \tag{14}$$

where

$$\Phi_t := \begin{bmatrix} \phi_t \\ \Psi_t \end{bmatrix} \in \mathbb{R}^n, \qquad \Theta^* := \begin{bmatrix} \vartheta^* \\ \theta^* \end{bmatrix} \in \mathbb{R}^n \, , \tag{15}$$

with $n = m + L$.

Although the system is nonlinear in the original input space, it becomes linear with respect to the unknown parameter vector Θ^* in the induced feature space, enabling the application of linear estimation methods for nonlinear system identification.

Under the bounded-noise assumption, each observation satisfies

$$\left| y_t - \Phi_t^\top \Theta^* \right| \leqslant \delta_e \, . \tag{16}$$

The set $\mathcal{S}_t^{\mathcal{H}} = \bigcap_{i=1}^{N} \mathcal{B}_i^{\mathcal{H}}$ of all parameter vectors compatible with the available observations defines the kernel-induced feasible region in the parameter space, where each kernel-induced strip $\mathcal{B}_i^{\mathcal{H}}$ is given by

$$\mathcal{B}_i^{\mathcal{H}} = \left\{ \Theta \in \mathbb{R}^n : \left| y_i - \Phi_i^\top \Theta \right| \leqslant \delta_e \right\} . \tag{17}$$

Geometrically, $\mathcal{B}_t^{\mathcal{H}}$ corresponds to a strip bounded by two parallel hyperplanes in the finite-dimensional parameter space $\mathbb{R}^n$. Although the underlying model describes a nonlinear system represented in an RKHS, the set-membership constraint is expressed in terms of the parameter vector Θ^*, which parameterizes the function f through a finite kernel expansion. Since this geometric structure is identical to that in the linear set-membership formulation, the intersection between $\mathcal{B}_t^{\mathcal{H}}$ and a previous ellipsoidal outer bound can be handled using the same ellipsoidal outer bounding in the linear regression. As a result, ellipsoidal outer bounding algorithm developed for linear models can be directly extended to the kernel-based formulation by replacing the linear regressor with its kernel-induced counterpart.

The objective is to estimate the unknown model parameters and the noise bound from the available input-output data $\{u_t, y_t\}_{t=1}^{N}$, while ensuring consistency with the kernel-induced feasible region $\mathcal{S}_t^{\mathcal{H}}$.

3.2. Kernel-based Ellipsoidal Outer-bounding Algorithm

The proposed algorithm aims to estimate the unknown parameter vector by recursively constructing ellipsoidal sets that contain all parameter vectors consistent with the available observations. At each iteration, the ellipsoid ξ_t is computed as an outer approximation of the intersection between the previous ellipsoid ξ_{t-1} and the kernel-induced strip $\mathcal{B}_t^{\mathcal{H}}$, such that

$$\xi_t \supseteq \left(\xi_{t-1} \cap \mathcal{B}_t^{\mathcal{H}} \right) . \tag{18}$$

The ellipsoidal outer bound at time $t-1$ is defined as

$$\xi_{t-1} = \left\{ \Theta \in \mathbb{R}^n \;\middle|\; \left(\Theta - \hat{\Theta}_{t-1}\right)^\top P_{t-1}^{-1} \left(\Theta - \hat{\Theta}_{t-1}\right) \leqslant \rho_{t-1} \right\} , \tag{19}$$

where P_{t-1} is a positive definite matrix characterizing the shape and orientation of the ellipsoid, $\hat{\Theta}_{t-1}$ denotes its center and the current parameter estimate, and $\rho_{t-1} > 0$ determines its size.

From Eqs. (18) and (19), the updated ellipsoid is computed as

$$\xi_t = \left\{ \Theta \in \mathbb{R}^n \;\middle|\; \begin{aligned} &\left(\Theta - \hat{\Theta}_{t-1}\right)^\top P_{t-1}^{-1} \left(\Theta - \hat{\Theta}_{t-1}\right) \\ &\qquad + \sigma_t \left(y_t - \Phi_t^\top \Theta\right)^2 \leqslant \rho_{t-1} + \sigma_t \, \delta_e^2 \end{aligned} \right\} . \tag{20}$$

The explicit expressions of the updated ellipsoidal parameters P_t, $\hat{\Theta}_t$, and ρ_t are given by [22]:

$$P_t = \left(I_n - \frac{G_t \, \sigma_t}{1 + G_t \, \sigma_t} \right) P_{t-1} \, , \tag{21}$$

where

$$G_t = \Phi_t^\top P_{t-1} \, \Phi_t \, . \tag{22}$$

The updated center of the ellipsoid is obtained as

$$\hat{\Theta}_t = \hat{\Theta}_{t-1} + \frac{P_{t-1} \Phi_t \, \sigma_t}{1 + G_t \sigma_t} \, \varepsilon_{t/\hat{\Theta}_{t-1}} \, , \tag{23}$$

and the corresponding size parameter is updated according to

$$\rho_t = \rho_{t-1} + \sigma_t \, \delta_e^2 - \frac{\sigma_t \, \varepsilon_{t/\hat{\Theta}_{t-1}}^2}{1 + G_t \sigma_t} \, . \tag{24}$$

The a priori estimation error is defined as

$$\varepsilon_{t/\hat{\Theta}_{t-1}} = y_t - \Phi_t^\top \hat{\Theta}_{t-1} \, . \tag{25}$$

The scalar parameter σ_t is computed to reduce the magnitude of the a posteriori estimation error below the noise bound δ_e, while minimizing the volume of the resulting ellipsoid

$$\mathcal{V}_{\xi_t} \triangleq \det(\rho_t P_t) \, . \tag{26}$$

The parameter σ_t is computed as

$$\sigma_t = \begin{cases} \dfrac{1}{G_t}\left(\dfrac{|\varepsilon_{t/\hat{\Theta}_{t-1}}|}{\delta_e} - 1\right), & \text{if } |\varepsilon_{t/\hat{\Theta}_{t-1}}| > \delta_e \text{ and } G_t > 0 \\ 0, & \text{if } a_t \geqslant 0 \\ \dfrac{-b_t + \sqrt{b_t^2 - 4a_t c_t}}{2a_t}, & \text{otherwise} \end{cases} \tag{27}$$

where

$$\begin{cases} a_t = \delta_e^2 (n-1)\, G_t^2 \\ b_t = \left((2n-1)\delta_e^2 - \rho_{t-1} G_t + \varepsilon_{t/\hat{\Theta}_{t-1}}^2\right) G_t \, . \\ c_t = n\left(\delta_e^2 - \varepsilon_{t/\hat{\Theta}_{t-1}}^2\right) - \rho_{t-1}\, G_t \end{cases} \tag{28}$$

According to the value of σ_t defined in Eq. (27), two cases can be distinguished. First, when the magnitude of the a priori prediction error satisfies

$$|\varepsilon_{t/\hat{\Theta}_{t-1}}| > \delta_e \, , \tag{29}$$

the algorithm focuses on reducing the estimation error. In this case, σ_t is computed to decrease the magnitude of the a priori prediction error below the noise bound δ_e, while simultaneously reducing the volume of the ellipsoid.

Conversely, when

$$\left|\varepsilon_{t/\hat{\Theta}_{t-1}}\right| \leqslant \delta_e \, , \tag{30}$$

the available observation is already consistent with the current ellipsoidal estimate. Therefore, the algorithm focuses on minimizing the volume of the ellipsoid while preserving the consistency condition with the a priori prediction error noise bound.

The proposed Algorithm 1 is referred as the kernel-based EOB.

The Theorem 3 describes the behavior of the a posteriori prediction error with respect to δ_e, as well as the evolution of the ellipsoid volume $\mathcal{V}_{\xi_t}$.

Theorem 3. Consider the Eq. (14) and the kernel-based EOB identification algorithm presented in Algorithm 1. Then, the following properties hold:

- $\underline{P_1}|$ If $G_t > 0$, then the a posteriori prediction error $\varepsilon_{t/\hat{\Theta}_t}$ satisfies

$$|\varepsilon_{t/\hat{\Theta}_t}| \leqslant \delta_e \, . \tag{31}$$

- $\underline{P_2}|$ The volume $\mathcal{V}_{\xi_t}$ of the ellipsoid ξ_t satisfies

$$\mathcal{V}_{\xi_t} \leqslant \mathcal{V}_{\xi_{t-1}} \, . \tag{32}$$

The proof of Theorem 3 is presented in Appendix A.

Algorithm 1 Kernel-based EOB algorithm

Input: $\{y_t, u_t\}_{t=1}^N$, $\hat{\Theta}_0$, P_0, ρ_0, δ_e, $\kappa(\cdot,\cdot)$, L
Output: $\hat{\Theta}_t$

1: Construct the dictionary $\mathcal{D} = \{\phi_1, \ldots, \phi_L\}$
2: **for** $t = 1$ to N **do**
3: Construct the kernel regressor

$$\Psi_t = \left[\kappa(\phi_1, \phi_t), \ldots, \kappa(\phi_L, \phi_t)\right]^\top .$$

4: Construct the augmented regressor

$$\Phi_t = \left[\phi_t^\top, \Psi_t^\top\right]^\top .$$

5: Compute $\hat{y}_{t/\hat{\Theta}_{t-1}}$, $\varepsilon_{t/\hat{\Theta}_{t-1}}$, and G_t.
6: Compute σ_t according to Eq. (27).
7: Update P_t according to Eq. (21).
8: Update $\hat{\Theta}_t$ according to Eq. (23).
9: Update ρ_t according to Eq. (24).
10: **end for**

The convergence behavior of the parameter vector estimated by the kernel-based EOB algorithm towards the true parameter vector is established in Theorem 4.

Theorem 4. Consider the Eq. (14) and the kernel-based EOB identification algorithm presented in Algorithm 1. If $P = p_0 I_n$, where $p_0 > 0$, and if the sequence $\left\{\sqrt{\sigma_t}\Phi_t\right\}_{t=1}^N$ satisfies the persistent excitation condition of order $o_m \geqslant n$, i.e., there exist two positive constants $\alpha > 0$ and $\beta > 0$ such that, for all $t \geqslant o_m$,

$$\alpha\, I_n \leqslant \sum_{j=0}^{o_m - 1} \Phi_{t-j}\, \sigma_{t-j} \Phi_{t-j}^\top \leqslant \beta I_n \, . \tag{33}$$

Then, for any initial condition such that $\Theta_0 \in \mathbb{R}^n$, and as long as $|\varepsilon_{t/\hat{\Theta}_{t-1}}| > \delta_e$, the following property holds:

- The parameter estimation error satisfies

$$\left\|\tilde{\Theta}_t\right\|_2^2 \leqslant \frac{1}{p_0 \left\lfloor \dfrac{t}{o_m} \right\rfloor \alpha + 1} \left\|\tilde{\Theta}_0\right\|_2^2 . \tag{34}$$

  where $\tilde{\Theta}_t = \Theta^* - \hat{\Theta}_t$ denotes the parameter error vector and $\lfloor \cdot \rfloor$ denotes the floor operator and $\| \cdot \|_2$ represents the Euclidean norm.

The proof of Theorem 4 is presented in Appendix B.

3.3. Adaptive Estimation of the Noise Bound

The kernel-based algorithm requires prior knowledge of an upper bound on the noise. In practice, however, this bound is generally unknown. An inappropriate choice of this bound may lead to either an underestimation or an overestimation, which can degrade the identification performance.

To overcome this limitation, an adaptive recursive procedure is proposed to estimate the unknown noise bound online. Unlike offline approaches, the proposed strategy updates the noise bound at each sampling instant using only the available measurements. Therefore, the data are processed sequentially from $t = 1$ to $t = N$.

Let $\hat{\delta}_t$ denote the estimated noise bound at time t. At each iteration, the kernel-based-EOB algorithm is executed using

the current noise bound estimate $\hat{\delta}_{t-1}$. The value of σ_t is then used to adaptively update $\hat{\delta}_t$ according to the following two cases.

1) If $\sigma_t \neq 0$, the current estimate of the noise bound is considered to be underestimated. Accordingly, the estimated bound is increased by a positive increment $\varepsilon_1 > 0$, according to
$$\hat{\delta}_t = \hat{\delta}_{t-1} + \varepsilon_1 \,. \tag{35}$$
where $\varepsilon_1 > 0$ is a small positive constant.
2) If $\sigma_t = 0$, the estimated bound is considered to be overestimated, i.e., $\hat{\delta}_t > \delta_e$. In this case, the estimated bound is recursively reduced. To this end, we propose to adopt a strategy similar to those presented in [21].

From Eq. (27), the condition $\sigma_t = 0$ is satisfied when

$$c_t = n\left(\left(\hat{\delta}_t\right)^2 - \varepsilon^2_{t/\hat{\Theta}_{t-1}}\right) - \rho_{t-1}\Phi_t^\top P_{t-1}\Phi_t \geqslant 0 \,. \tag{36}$$

Consequently,

$$\left(\hat{\delta}_t\right)^2 \geqslant \varepsilon^2_{t/\hat{\Theta}_{t-1}} + \frac{\rho_{t-1}\Phi_t^\top P_{t-1}\Phi_t}{n} \,. \tag{37}$$

The index corresponding to the largest a priori prediction error is defined as

$$J = \arg\max_{1\leqslant i\leqslant t} \left|\varepsilon_{i/\hat{\Theta}_{i-1}}\right| \,. \tag{38}$$

By definition of J

$$\varepsilon^2_{J/\hat{\Theta}_{J-1}} \geqslant \varepsilon^2_{i/\hat{\Theta}_{i-1}}, \qquad i = 1, \ldots, t \,. \tag{39}$$

Evaluating Eq. (37) at $t = J$ gives

$$\left(\hat{\delta}_J\right)^2 \geqslant \varepsilon^2_{J/\hat{\Theta}_{J-1}} + \frac{\rho_{J-1}\Phi_J^\top P_{J-1}\Phi_J}{n} \,. \tag{40}$$

The following theorem provides a probabilistic lower bound on the maximum a priori prediction error generated during the recursive estimation process.

Theorem 5. Assume that the noise sequence $\{e_t\}$ is independent and identically distributed (i.i.d.). Let $\varepsilon_2 > 0$ be an arbitrarily small constant and define

$$\mathcal{B}^+_{\varepsilon_2} = [\delta_e - \varepsilon_2, \delta_e] \,, \qquad \mathcal{B}^-_{\varepsilon_2} = [-\delta_e, -\delta_e + \varepsilon_2] \,. \tag{41}$$

If there exists a constant $\Delta > 0$ such that

$$P(e_t \in \mathcal{B}^+_{\varepsilon_2}) > \Delta \,, \qquad P(e_t \in \mathcal{B}^-_{\varepsilon_2}) > \Delta \,, \qquad t = 1, \ldots, N \,. \tag{42}$$

then, with probability at least $1 - (1-\Delta)^N$ there exists an index J satisfying

$$\varepsilon^2_{J/\hat{\Theta}_{J-1}} \geqslant (\delta_e - \varepsilon_2)^2. \tag{43}$$

The proof of Theorem 5 is presented in Appendix C.

The i.i.d. assumption introduced in Theorem 5 is considered solely for the theoretical analysis of the proposed noise-bound estimation procedure.

The identification algorithm itself remains formulated within the set-membership identification framework and does not rely on any specific probabilistic model of the disturbance. The proposed procedure is applied to the example systems presented in Sec. 4.1; however, the validity of this theoretical assumption cannot be guaranteed a priori.

Combining (40) with Theorem 5 leads to

$$\delta_e^2 \leqslant \left(\hat{\delta}_J\right)^2 - \frac{\rho_{J-1}\Phi_J^\top P_{J-1}\Phi_J}{n} + \varepsilon_2(2\delta_e - \varepsilon_2) \,, \tag{44}$$

with probability at least $1 - (1-\Delta)^N$.

Under the assumption $\hat{\delta}_{J-1} \geqslant \delta_e$, the recursive reduction is defined as

$$\delta_e^2 \leqslant \left(\hat{\delta}_J\right)^2 - d_J \,, \tag{45}$$

where

$$d_J = \frac{\rho_{J-1}\Phi_J^\top P_{J-1}\Phi_J}{n} - \varepsilon_2\left(2\hat{\delta}_{J-1} - \varepsilon_2\right) . \tag{46}$$

The adaptive update is finally given by

$$\hat{\delta}_t = \begin{cases} \sqrt{(\hat{\delta}_{t-1})^2 - d_J}, & d_J > 0 \\ \hat{\delta}_{t-1}, & d_J \leqslant 0 \end{cases} . \tag{47}$$

The complete kernel-based EOB algorithm with recursive automatic noise-bound estimation is summarized as Algorithm 2.

Algorithm 2 Kernel-based EOB with recursive automatic noise-bound estimation

Input: $\{y_t, u_t\}_{t=1}^N$, $\hat{\Theta}_0$, P_0, ρ_0, $\hat{\delta}_0$, $\kappa(\cdot,\cdot)$
Output: $\hat{\delta}_t$, $\hat{\Theta}_t$

for $t = 1$ to N **do**
  Execute kernel-based EOB algorithm using $\hat{\delta}_{t-1}$.
  if $\sigma_t \neq 0$ **then**
    Update $\hat{\delta}_t$ according to Eq. (35).
  else
    Compute
$$J = \arg\max_{1\leqslant i\leqslant t} |\varepsilon_{i/\hat{\Theta}_{i-1}}|.$$
    Compute d_J according to Eq. (46).
    Update $\hat{\delta}_t$ according to Eq. (47).
  end if
end for
return $\hat{\delta}_t$ and $\hat{\Theta}_t$

4. Numerical and Experimental Results

4.1. Simulation Results

The first benchmark experiment considers the following nonlinear system [6]:

$$\begin{cases} z_t = 0.6\, z_{t-1} + 0.2\, z_{t-2} + 0.3 \exp(u_{t-1}) \\ y_t = z_t^2 + e_t \end{cases} . \tag{48}$$

The input signal u_t is assumed to be a stationary colored Gaussian process generated by a second-order autoregressive

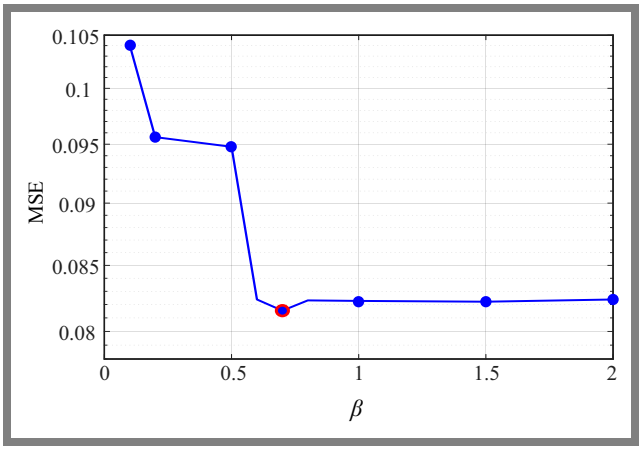


Fig. 2. Performance comparison of the proposed algorithm for different values of β.

AR(2) model with poles located at $(0.3, 0.8)$. The noise sequence e_t is bounded by the upper and lower bound $\pm\delta_e$, where $\delta_e = 0.5$.

The dataset consists of $N = 4000$ samples, which are divided into two separate subsets. The first 2000 samples are used for training the algorithms, whereas the remaining 2000 samples are employed for evaluating their performance. This data partition ensures that the algorithms are assessed using unseen samples.

4.2. First Experiment

In first experiment, the objective is to evaluate the performance of the proposed algorithm implemented using a Gaussian kernel. The kernel parameter is set to $\sigma = 0.7$, as determined through the performance analysis presented in Fig. 2. In addition, the dictionary size is fixed at $L = 80$, based on the analysis reported in Fig. 3.

The noise bound estimation process is performed using the following synthesis parameters: $\varepsilon_1 = 10^{-5}$, $\varepsilon_2 = 0.25$, and $\hat{\delta}_0 = 2$.

Figure 2 presents the variation of the mean squared error (MSE) with respect to the kernel parameter β. It can be observed that the MSE decreases rapidly as β increases from 0.1 to 0.7, indicating a significant improvement in identification accuracy. The minimum MSE is achieved at $\beta = 0.7$. Beyond this value, the MSE remains nearly constant, suggesting that further increasing β does not provide any performance improvement. Consequently, $\beta = 0.7$ is adopted in all simulations.

The influence of the dictionary size on the identification performance is illustrated in Fig. 3. As the dictionary size increases from $L = 10$ to $L = 80$, the MSE decreases progressively, indicating that a larger dictionary improves the approximation capability of the proposed algorithm. The minimum MSE is achieved at $L = 80$. However, further increasing the dictionary size results in a slight degradation of the identification performance, as evidenced by the increase in the MSE for $L > 80$. This behavior suggests that enlarging the dictionary beyond $L = 80$ introduces redundant basis functions without improving the model accuracy. Therefore,

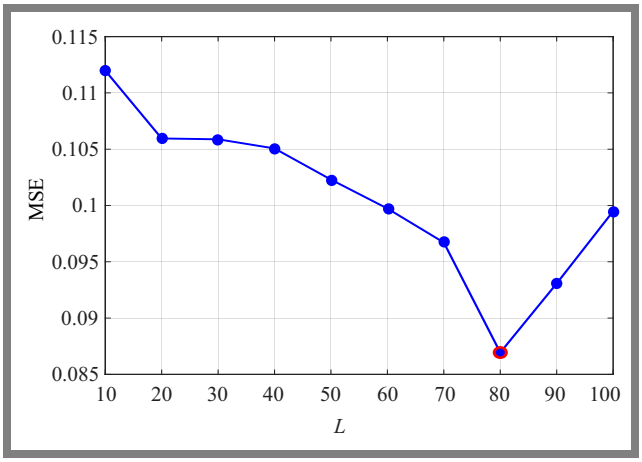


Fig. 3. Performance comparison of the proposed algorithm for different values of the dictionary size L.

the dictionary size is fixed at $L = 80$ for all subsequent simulations, providing the best identification performance.

Figure 4 illustrates the convergence of the estimated noise bound $\hat{\delta}_t$ toward the true noise bound δ_e as a function of time t. This validates the automatic noise bound estimation algorithm proposed in Subsection 3.3. Figure 5 presents the evolution of the a posteriori estimation error $\varepsilon_{t/\hat{\Theta}_{t-1}}$ along with the estimated noise bound $\hat{\delta}_t$ as a function of time t. It can be observed that the estimation error remains bounded by the estimated noise bound. This result confirms the theoretical guarantee established in Theorem 3. Figure 6 illustrates the true validation output y_t and the corresponding estimated output $\hat{y}_t$ obtained on the validation dataset using the proposed algorithm, thereby confirming Theorem 4.

To further evaluate the effectiveness and robustness of the proposed algorithm, additional experiment were conducted under different noise bounds, $\delta_e = 0.1, 0.5$, and 1.

Figure 7 presents the evolution of the MSE as a function of time t for different values of the noise bound δ_e. It can be observed that the MSE level depends on the considered noise bound. More specifically, the MSE obtained for $\delta_e = 0.1$ is lower than those obtained for $\delta_e = 0.5$ and $\delta_e = 1$. Thus, increasing the noise bound δ_e results in a higher MSE level.

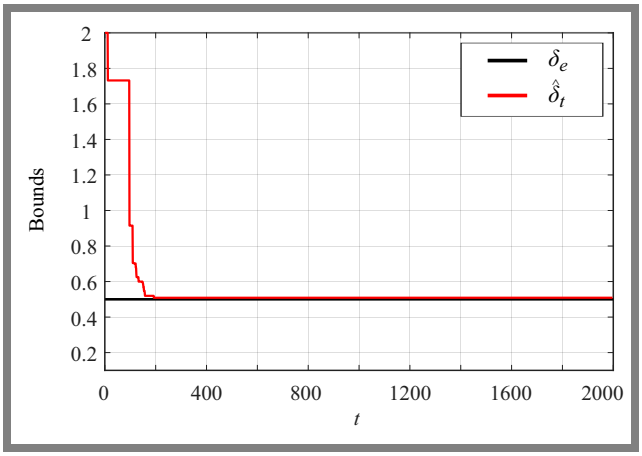


Fig. 4. Convergence of the estimated bound $\hat{\delta}_t$ toward the true bound δ_e.

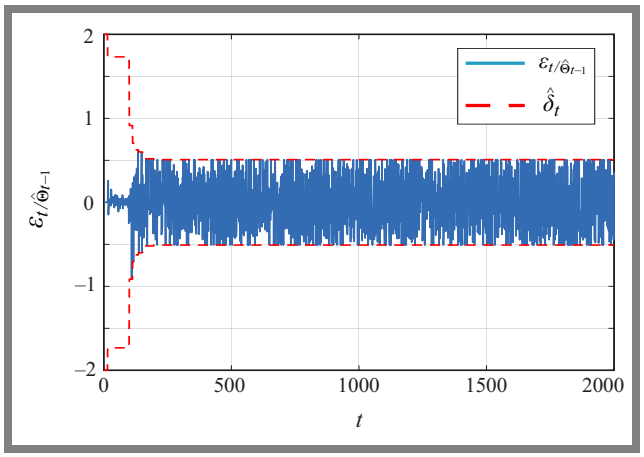


Fig. 5. A posteriori error $\varepsilon_{t/\hat{\Theta}_{t-1}}$ and the estimated noise bound $\hat{\delta}_t$.

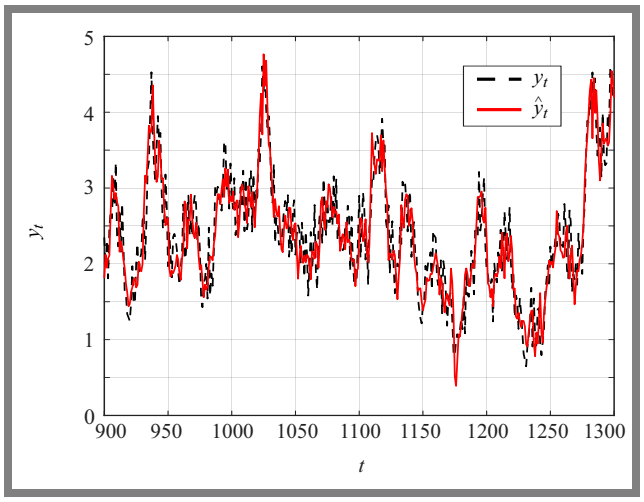


Fig. 6. True output y_t and estimated output $\hat{y}_t$.

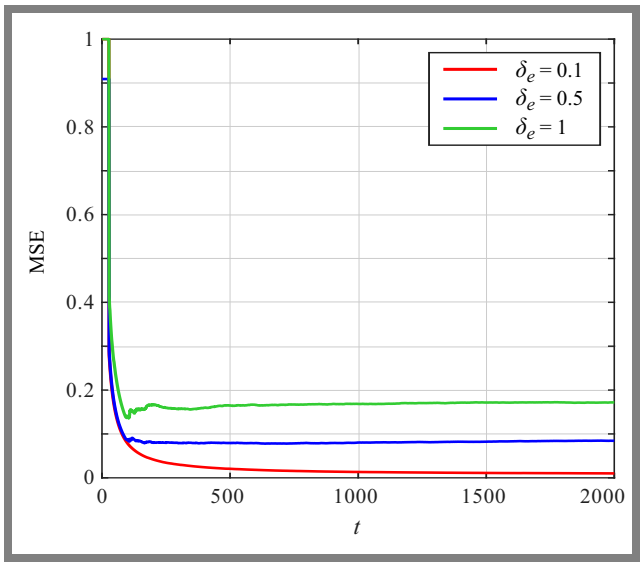


Fig. 7. Performance comparison of the proposed algorithm for different values of δ_e.

To further investigate the robustness of the proposed algorithm with respect to the statistical characteristics of the disturbance, three noise sources with different compositions are considered:

1) $e_t = e_t^{(1)} = \delta_e \left(\frac{9}{10} n_t + \frac{1}{10} b_t \right)$,

2) $e_t = e_t^{(2)} = \delta_e \left(\frac{5}{10} n_t + \frac{5}{10} b_t \right)$,

3) $e_t = e_t^{(3)} = \delta_e \left(\frac{1}{10} n_t + \frac{9}{10} b_t \right)$,

where n_t denotes a white noise sequence uniformly distributed over $[-1, 1]$, while b_t denotes a binary square wave signal with a period of two samples. By varying the relative contributions of n_t and b_t, these three disturbance signals allow the behavior of the proposed algorithm to be assessed under different noise characteristics.

Tab. 2. Performance comparison under different noise types.

Noise	RMSE	MAE	VAF [%]	NBV [%]
e_t^1	0.08325	0.058736	99.150	0.955
e_t^2	0.18344	0.147270	97.201	1.160
e_t^3	0.27885	0.230340	91.988	1.920

The root mean square error (RMSE), mean absolute error (MAE), variance accounted for (VAF), and number of bound violations (NBV) were adopted as performance criteria. These criteria are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} \left(y_t - \hat{y}_t \right)^2}, \tag{49}$$

$$\text{MAE} = \frac{1}{N} \sum_{t=1}^{N} |y_t - \hat{y}_t| , \tag{50}$$

$$\text{VAF} = \left(1 - \frac{\text{Var}(y_t - \hat{y}_t)}{\text{Var}(y_t)} \right) \times 100 , \tag{51}$$

$$\text{NBV} = \frac{1}{N} \sum_{t=1}^{N} \mathbf{1} \left(|y_t - \hat{y}_t| > \hat{\delta}_e \right) \times 100 , \tag{52}$$

where $\mathbf{1}(\cdot)$ denotes the indicator function defined as:

$$\mathbf{1}(A) = \begin{cases} 1, & \text{if } A \text{ is true} \\ 0, & \text{otherwise} \end{cases} . \tag{53}$$

The results reported in Tab. 2 show that, for the three considered noise distributions, the proposed algorithm achieves high VAF values, together with relatively low RMSE, MAE, and NBV. These results confirm that the proposed algorithm maintains good performance across different noise distributions and does not require the noise to follow a specific distribution, demonstrating its robustness with respect to the noise characteristics.

Figure 8 illustrates the convergence of the estimated bounds $\hat{\delta}_t$ toward the true noise bound δ_e for the three considered noise types. It can be observed that all estimated bounds for the three considered noise types converge toward the true

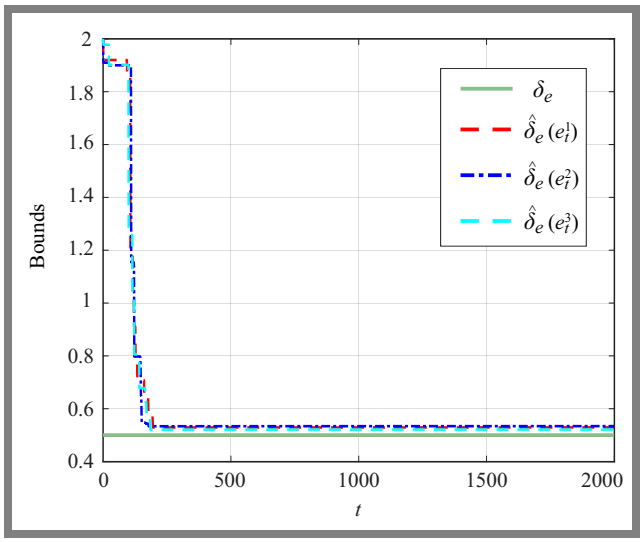


Fig. 8. Convergence of the estimated bounds $\hat{\delta}_t$ toward the true noise bound δ_e for the three considered noise types.

Tab. 3. Parameter settings of the considered algorithms.

Algorithm	β	$\varepsilon_1, \varepsilon_2$	δ_0	Complexity
Kernel-based EOB	0.7	$\varepsilon_1 = 10^{-5}$, $\varepsilon_2 = 0.25$	2	$\mathcal{O}(L^2)$
KEOB	0.5	–	-	$\mathcal{O}(L^2)$
SM-KRLS	5.1	–	-	$\mathcal{O}(L^2)$

bound δ_e, confirming the effectiveness of the proposed bound estimation procedure.

4.3. Second Experiment

The objective of the second experiment is to compare the proposed algorithm with existing ellipsoidal algorithms for nonlinear systems: KEOB [20] and SM-KRLS [6]. For a fair comparison, the parameter settings reported in [20] and [6] are adopted for the corresponding methods. The dictionary size is fixed to $L = 80$ for the proposed kernel-based EOB and KEOB, whereas the dictionary construction is used for SM-KRLS. The same data, noise conditions, and performance metrics as those used in the first experiment are adopted for all considered methods to ensure a consistent and fair comparison. The resulting hyperparameter settings are summarized in Tab. 3.

To evaluate and compare the performance of the three algorithms, 100 independent Monte Carlo simulations were conducted under identical experimental conditions.

Figure 9 illustrates the evolution of the MSE computed on the validation data over time. It can be observed that the proposed algorithm achieves the lowest MSE compared to the KEOB and SM-KRLS algorithms.

Table 4 summarizes the performance comparison between the proposed algorithm and the KEOB and SM-KRLS algorithms. The results demonstrate that the proposed algorithm achieves the best overall performance across all evaluation metrics. Specifically, KB-EOB provides the highest FIT and VAF

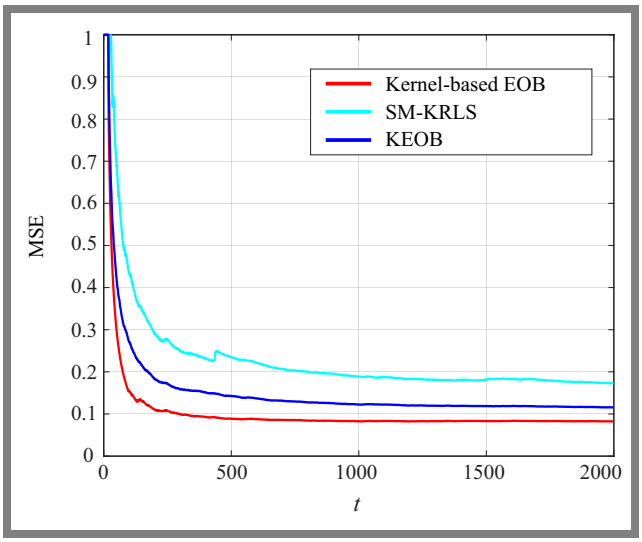


Fig. 9. Evolution of the MSE over time t.

Tab. 4. Performance comparison of the considered identification algorithms.

Metric	Kernel-based EOB	KEOB	SM-KRLS
FIT [%]	**91.731**	88.590	76.263
RMSE	**0.2871**	0.3400	0.4156
MAE	**0.1986**	0.2468	0.3185
VAF [%]	**94.529**	91.313	84.812
NBV [%]	**1.6633**	3.840	27.663

values, reaching 91.731% and 94.529%, respectively, while also achieving the lowest RMSE, MAE, and NBV values.

Figure 10 illustrates the evolution of the logarithmic volume ratio $\log_{10}(\mathcal{V}_{\xi_t}/\mathcal{V}_{\xi_0})$ over time for the proposed algorithm and the KEOB. For the SM-KRLS algorithm, the size of the ellipsoid is not available in [6]. The logarithmic volume ratio decreases progressively, indicating a continuous reduction in the ellipsoid volume. Moreover, the proposed algorithm achieves a smaller ellipsoid volume than KEOB, resulting in a tighter parameter uncertainty set.

4.4. Third Experiment

In the third experiment, we focus on the identification of a model for nitrogen dioxide (NO_2) concentration in Paris, along the section between Porte de la Chapelle and Porte de Bagnolet (Chap-Bagn), using the proposed algorithm. Due to the complex and nonlinear interactions between air-quality and meteorological variables, the data are considered within a nonlinear system identification framework.

The data used in this experiment are obtained from open-access air-quality datasets provided by the Air Quality Observatory in the Île-de-France region.[1]

A total of $N = 7 \times 10^3$ samples are available for this experiment and are divided into two distinct datasets. The first 3.5×10^3 samples are used for training the algorithm,

[1] https://data-airparif-asso.opendata.arcgis.com

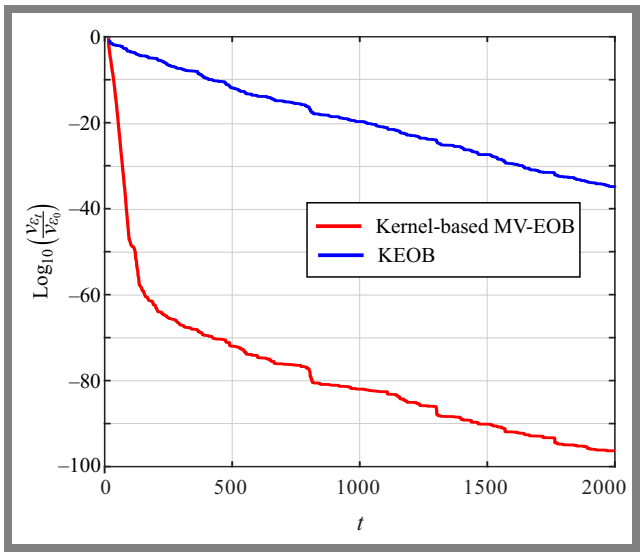


Fig. 10. Logarithmic volume ratio $\log_{10}\left(\mathcal{V}_{\xi_t}/\mathcal{V}_{\xi_0}\right)$ as a function of t.

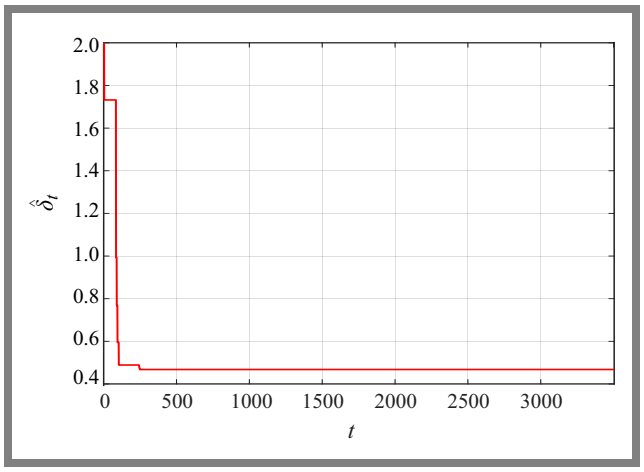


Fig. 11. The estimated bound $\hat{\delta}_t$ as a function of t.

while the remaining 3.5×10^3 samples are used to validate its performance.

The value of β is set to 0.7, with a dictionary size of $L = 80$. The noise bound estimation process is performed using the following synthesis parameters: $\varepsilon_1 = 10^{-5}$, $\varepsilon_2 = 0.25$, and $\hat{\delta}_0 = 2$.

Figure 13 presents the converegence of the estimated noise bound. Figure 12 shows that the estimated concentration (red line) closely converges to the observed concentration (black line) over the validation data. Figure 10 illustrates the evolution of the logarithmic volume ratio $\log_{10}\left(\mathcal{V}_{\xi_t}/\mathcal{V}_{\xi_0}\right)$ over time.

5. Conclusions

This paper proposed a recursive kernel-based ellipsoidal outer-bounding algorithm for nonlinear system identification under unknown but bounded disturbances. By combining kernel learning with the set-membership identification framework, the nonlinear identification problem was transformed into a linear regression problem in a reproducing kernel Hilbert space, enabling recursive parameter estimation with guaranteed bounded-error properties.

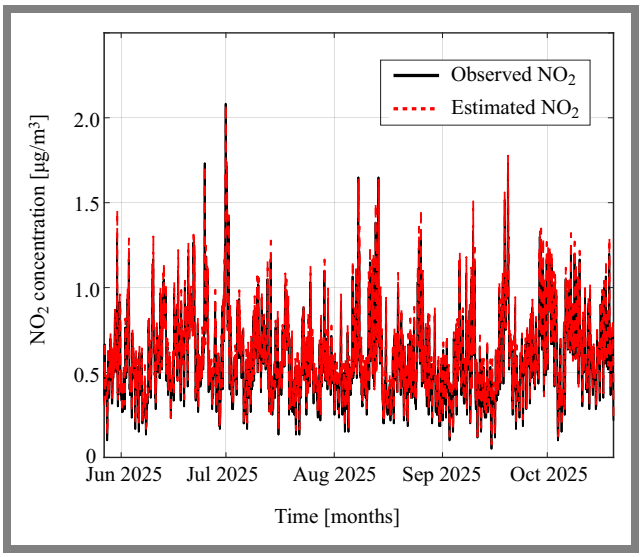


Fig. 12. Concentration of NO_2 from January to October 2025: process output (black line) versus estimated output (red line).

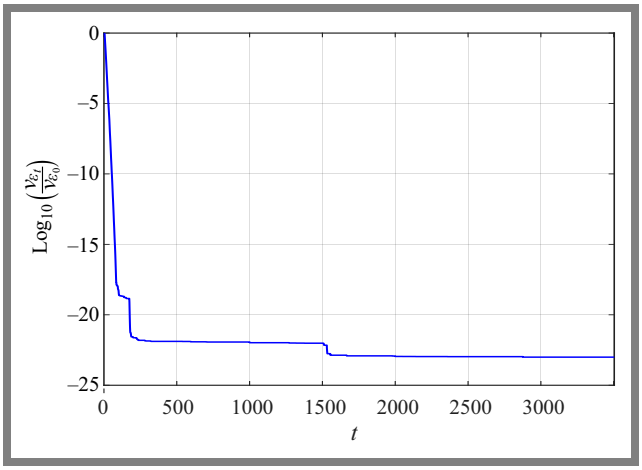


Fig. 13. Logarithmic volume ratio $\log_{10}\left(\mathcal{V}_{\xi_t}/\mathcal{V}_{\xi_0}\right)$ as a function of t.

A recursive data-driven procedure was introduced to jointly estimate the unknown disturbance bound and the model parameters, eliminating the need for prior knowledge of the noise level. Furthermore, a minimum volume ellipsoidal uncertainty region was recursively constructed to provide a compact representation of the feasible parameter set. The convergence properties of the proposed algorithm were also established theoretically.

Simulation results on nonlinear benchmark systems demonstrate the effectiveness of the proposed algorithm. Compared with existing kernel-based ellipsoidal algorithms, the proposed method achieves improved identification accuracy, more accurate disturbance-bound estimation, and fewer bound violations. Future work will focus on developing an adaptive strategy for improving dictionary selection.

Appendix A: Proof of Theorem 3

Proof. To prove $\underline{P_1}|$, we first premultiply Eq. (23) by $\Phi_t^\top$ and substitute the resulting expression into Eq. (25), which gives

$$\varepsilon_{t/\hat{\Theta}_t} = \frac{1}{1+G_t\sigma_t}\varepsilon_{t/\hat{\Theta}_{t-1}} . \tag{54}$$

The result is established by considering the following two cases.

1) If $|\varepsilon_{t/\hat{\theta}_{t-1}}| > \delta_e$, using Eq. (27) in Eq. (54) yields

$$|\varepsilon_{t/\hat{\Theta}_t}| = |\varepsilon_{t/\hat{\Theta}_{t-1}}| \left| \frac{\delta_e}{\varepsilon_{t/\hat{\Theta}_{t-1}}} \right| = \delta_e . \tag{55}$$

2) Otherwise, if $|\varepsilon_{t/\hat{\theta}_{t-1}}| \leqslant \delta_e$, it follows from Eq. (27) that $\sigma_t \geqslant 0$, and hence

$$0 < \frac{1}{1+G_t\,\sigma_t} \leqslant 1 . \tag{56}$$

Substituting this result into Eq. (54) gives

$$|\varepsilon_{t/\hat{\Theta}_t}| \leqslant |\varepsilon_{t/\hat{\Theta}_{t-1}}| \leqslant \delta_e . \tag{57}$$

Therefore, in both cases, $|\varepsilon_{t/\hat{\theta}_t}| \leqslant \delta_e$, which proves $\underline{P_1}|$.

To prove property $\underline{P_2}$, three cases are considered.

1) If $|\varepsilon_{t/\hat{\theta}_{t-1}}| > \delta_e$, then Eq. (27) implies $\sigma_t \geqslant 0$.
   Applying the matrix inversion lemma to Eq. (27) gives

$$P_t^{-1} = P_{t-1}^{-1} + \Phi_t\,\sigma_t\,\Phi_t^\top , \tag{58}$$

   which yields $P_t \leqslant P_{t-1}$.
   Moreover,

$$\rho_t = \rho_{t-1} + \sigma_t\left(\delta_e^2 - |\varepsilon_{t/\hat{\Theta}_{t-1}}||\varepsilon_{t/\hat{\Theta}_t}|\right) . \tag{59}$$

   Since $|\varepsilon_{t/\hat{\theta}_t}| = \delta$, we obtain $\rho_t < \rho_{t-1}$, and consequently, $\mathcal{V}_{\xi_t} < \mathcal{V}_{\xi_{t-1}}$.
2) If $|\varepsilon_{t/\hat{\theta}_{t-1}}| \leqslant \delta$ and $c_t < 0$, the volume ratio is expressed as

$$\mathcal{V}_{\xi_t} = \mathcal{H}(\sigma_t)\,\mathcal{V}_{\xi_{t-1}}, \tag{60}$$

   where

$$\mathcal{H}(\sigma_t) = \frac{1}{1+G_t\sigma_t}\left(\frac{\rho_{t-1}+\sigma_t\delta_e^2 - \frac{\sigma_t\varepsilon^2_{t/\hat{\Theta}_{t-1}}}{1+G_t\sigma_t}}{\rho_{t-1}}\right)^n . \tag{61}$$

   Since

$$\left.\frac{d\mathcal{H}(\sigma)}{d\sigma}\right|_{\sigma=0} = c_t < 0 , \qquad \lim_{\sigma\to+\infty}\frac{d\mathcal{H}(\sigma)}{d\sigma} = +\infty . \tag{62}$$

   $\mathcal{H}(\sigma)$ admits a unique minimizer σ_t. Therefore, $\mathcal{H}(\sigma_t) < 1$, which gives $\mathcal{V}_{\xi_t} < \mathcal{V}_{\xi_{t-1}}$.
3) If $|\varepsilon_{t/\hat{\theta}_{t-1}}| \leqslant \delta_e$ and $c_t \geqslant 0$, then $\sigma_t = 0$. Hence

$$P_t = P_{t-1} , \qquad \rho_t = \rho_{t-1} , \tag{63}$$

   and therefore, $\mathcal{V}_{\xi_t} = \mathcal{V}_{\xi_{t-1}}$.

This completes the proof. □

Appendix B: Proof of Theorem 4

Proof. First, consider the Lyapunov function

$$\mathbf{V}_t = \tilde{\Theta}_t^\top P_t^{-1}\tilde{\Theta}_t . \tag{64}$$

Using the parameter update equation and the expression of the parameter error, we obtain

$$\mathbf{V}_t = \mathbf{V}_{t-1} + \mathbf{Q}_t , \tag{65}$$

with

$$\mathbf{Q}_t = \tilde{\Theta}_t^\top\hat{\Phi}_t\sigma_t\Phi_t^\top\tilde{\Theta}_t - 2\tilde{\Theta}_t^\top\Phi_t\sigma_t\varepsilon_{t/\hat{\Theta}_t} - G_t\sigma_t^2\varepsilon^2_{t/\hat{\Theta}_t} . \tag{66}$$

where from Eqs. (54) and (27),

$$\mathbf{Q}_t \leqslant \sigma_t\left(\left(\varepsilon_{t/\hat{\Theta}_t} - \hat{\Phi}_t^\top\tilde{\Theta}_t\right)^2 - \delta_e^2\right) \leqslant 0 . \tag{67}$$

Hence $\mathbf{V}_t \leqslant \mathbf{V}_{t-1}$ which implies

$$\tilde{\Theta}_t^\top P_t^{-1}\tilde{\Theta}_t \leqslant \tilde{\Theta}_0^\top P_0^{-1}\tilde{\Theta}_0 . \tag{68}$$

Moreover, since

$$P_t^{-1} = P_{t-1}^{-1} + \hat{\Phi}_t\sigma_t\hat{\Phi}_t^\top , \tag{69}$$

the persistent excitation condition yields

$$P_t^{-1} \geqslant P_0^{-1} + \left\lfloor\frac{t}{o_m}\right\rfloor\alpha I_n . \tag{70}$$

Combining the above inequalities establishes the boundedness and convergence properties of the parameter estimation error stated in the theorem. □

Appendix C: Proof of Theorem 5

Proof. From Eq. (14) and the definition of the parameter estimation error, the a posteriori prediction error can be expressed as

$$\varepsilon_{t/\hat{\Theta}_t} = \Phi_t^\top\tilde{\Theta}_t + e_t , \tag{71}$$

To characterize the cases where the prediction error reaches the noise bound, we introduce the following events:

$$E_t^- = \left\{e_t \in \mathcal{I}^-_{\varepsilon_2},\ \Phi_t^\top\tilde{\Theta}_t \leqslant 0\right\}, \tag{72}$$

$$E_t^+ = \left\{e_t \in \mathcal{I}^+_{\varepsilon_2},\ \Phi_t^\top\tilde{\Theta}_t \geqslant 0\right\}. \tag{73}$$

Let

$$E_t = E_t^- \cup E_t^+,$$

and denote its complementary event by E_t^c.

Since the noise sequence $\{e_t\}$ is assumed to be independent and identically distributed, e_t is independent of the estimation error term $\Phi_t^\top\tilde{\Theta}_t$. Moreover, for sufficiently small ε_2, the events E_t^- and E_t^+ are mutually exclusive. Therefore

$$P(E_t) = P(E_t^-) + P(E_t^+) . \tag{74}$$

Using the independence assumption gives

$$P(E_t) = P\left(e_t \in \mathcal{I}^-_{\varepsilon_2}\right) P\left(\Phi_t^\top \tilde{\Theta}_t \leqslant 0\right) + P\left(e_t \in \mathcal{I}^+_{\varepsilon_2}\right) P\left(\Phi_t^\top \tilde{\Theta}_t \geqslant 0\right). \tag{75}$$

According to Eq. (42)

$$P\left(e_t \in \mathcal{I}^-_{\varepsilon_2}\right) \geqslant \Delta, \qquad P\left(e_t \in \mathcal{I}^+_{\varepsilon_2}\right) \geqslant \Delta\,. \tag{76}$$

Hence

$$P(E_t) \geqslant \Delta P\left(\Phi_t^\top \tilde{\Theta}_t \leqslant 0\right) + \Delta P\left(\Phi_t^\top \tilde{\Theta}_t \geqslant 0\right) = \Delta\,, \tag{77}$$

which immediately implies

$$P(E_t^c) \leqslant 1 - \Delta\,. \tag{78}$$

We next evaluate the probability that none of these events occurs during the N observations.

By the chain rule of conditional probabilities

$$P\left(\bigcap_{t=1}^{N} E_t^c\right) = \prod_{t=1}^{N} P\left(E_t^c \,\middle|\, \bigcap_{k=1}^{t-1} E_k^c\right)\,. \tag{79}$$

Since the noise sample e_t is independent of all previous observations and events

$$P\left(E_t^c \,\middle|\, \bigcap_{k=1}^{t-1} E_k^c\right) \leqslant 1 - \Delta\,. \tag{80}$$

Therefore

$$P\left(\bigcap_{t=1}^{N} E_t^c\right) \leqslant (1-\Delta)^N\,. \tag{81}$$

Taking the complementary event yields

$$P\left(\bigcup_{t=1}^{N} E_t\right) = 1 - P\left(\bigcap_{t=1}^{N} E_t^c\right) \geqslant 1 - (1-\Delta)^N\,. \tag{82}$$

Consequently, with probability at least $1-(1-\Delta)^N$, there exists an instant $t \in \{1,\ldots,N\}$ for which E_t occurs. By the definition of E_t, the prediction error necessarily satisfies

$$\varepsilon^2_{t/\hat{\Theta}_t} \geqslant (\delta_e - \varepsilon_2)^2\,, \tag{83}$$

which completes the proof.

□

Acknowledgments

The authors would like to thank ENSICAEN, the University of Caen Normandy, and Sultan Moulay Slimane University for providing the research environment and support that made this work possible.

References

[1] F. Li, T. Zheng, and Q. Cao, "Modeling and Identification for Practical Nonlinear Process Using Neural Fuzzy Network-based Hammerstein System", *Transactions of the Institute of Measurement and Control*, vol. 45, pp. 2091–2102, 2023 (https://doi.org/10.1177/01423312221143777).

[2] F. Li, L. Song, T. Wang, and R. Liu, "Parameter Identification for Nonlinear Hammerstein Models with Stacked Sparse Autoencoder Network", *Engineering Applications of Artificial Intelligence*, vol. 163, art. no. 113002, 2026 (https://doi.org/10.1016/j.engappai.2025.113002).

[3] A. Bouklata *et al.*, "Online Identification Method for Wiener Nonlinear Systems", *International Journal of Dynamics and Control*, vol. 13, art. no. 65, 2025 (https://doi.org/10.1007/s40435-024-01569-3).

[4] F. Li, S. Qian, N. He, and B. Li, "Estimation of Wiener Nonlinear Systems with Measurement Noises Utilizing Correlation Analysis and Kalman Filter", *International Journal of Robust and Nonlinear Control*, vol. 34, pp. 4706–4718, 2024 (https://doi.org/10.1002/rnc.7224).

[5] I. Constantin, J. Constantin, and A. Bigand, "A New Kernel RLS Algorithm for Systems with Bounded Noise", *International Journal of Adaptive Control and Signal Processing*, vol. 32, pp. 344–361, 2017 (https://doi.org/10.1002/acs.2845).

[6] R. Fateh, A. Darif, A. Boumezzough, S. Safi, and M. Frikel, "A Novel Kernel Algorithm for Finite Impulse Response Channel Identification", *Journal of Telecommunications and Information Technology*, vol. 92, pp. 84–93, 2023 (https://doi.org/10.26636/jtit.2023.169823).

[7] R. Fateh, A. Darif, and S. Safi, "Kernel and Linear Adaptive Methods for the BRAN Channels Identification", *Advanced Intelligent Systems for Sustainable Development (AI2SD'2020)*, vol. 1418, pp. 579–591, 2022 (https://doi.org/10.1007/978-3-030-90639-9_47).

[8] N. Aronszajn, "Theory of Reproducing Kernels", *Transactions of the American Mathematical Society*, vol. 68, pp. 337–404, 1950 (https://doi.org/10.1090/S0002-9947-1950-0051437-7).

[9] Y. Engel, S. Mannor, and R. Meir, "The Kernel Recursive Least-squares Algorithm", *IEEE Transactions on Signal Processing*, vol. 52, pp. 2275–2285, 2004 (https://doi.org/10.1109/TSP.2004.830985).

[10] B. Schölkopf, A. Smola, and K.R. Müller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem", *Neural Computation*, vol. 10, pp. 1299–1319, 1998 (https://doi.org/10.1162/089976698300017467).

[11] M. Zhang, X. Wang, X. Chen, and A. Zhang, "The Kernel Conjugate Gradient Algorithms", *IEEE Transactions on Signal Processing*, vol. 66, pp. 4377–4387, 2018 (https://doi.org/10.1109/TSP.2018.2853109).

[12] S. Zhao, B. Chen, and J.C. Principe, "Kernel Adaptive Filtering with Maximum Correntropy Criterion", *International Joint Conference on Neural Networks (IJCNN)*, San Jose, USA, 2011 (https://doi.org/10.1109/IJCNN.2011.6033473).

[13] J. Zhao, H. Zhang, and J.A. Zhang, "Gaussian Kernel Adaptive Filters with Adaptive Kernel Bandwidth", *Signal Processing*, vol. 166, art. no. 107270, 2020 (https://doi.org/10.1016/j.sigpro.2019.107270).

[14] Y. Li *et al.*, "Learning the Uncertainty Sets of Linear Control Systems via Set Membership: A Non-asymptotic Analysis", *Forty-First International Conference on Machine Learning*, Vienna, Austria, 2024 (https://proceedings.mlr.press/v235/li24ci.html).

[15] D. Chen, Z. Zhou, J. Hu, and J. Liu, "Set-membership State Estimation for Delayed Switched Systems with Fading Measurement and Interval Uncertainty", *Optimal Control Applications and Methods*, vol. 45, pp. 1691–1715, 2024 (https://doi.org/10.1002/oca.3118).

[16] M. Pouliquen, E. Pigeon, and O. Gehan, "Output Error Identification for Multi-input Multi-output Systems with Bounded Disturbances", *50th IEEE Conference on Decision and Control and European Control Conference*, Orlando, Florida, 2011 (https://doi.org/10.1109/CDC.2011.6160211).

[17] Y. Liu, Y. Zhao, and F. Wu, "Ellipsoidal State-bounding-based Set-membership Estimation for Linear Systems with Unknown-but-bounded Disturbances", *IET Control Theory & Applications*, vol. 10, pp. 431–442, 2016 (https://doi.org/10.1049/iet-cta.2015.0654).

[18] H. El Maizi, M. Pouliquen, S. Safi, and M. Frikel, "Identification of Output Error Model with Bounded Disturbances", *11th IEEE International Conference on Systems and Control (ICSC)*, Sousse, Tunisia, 2023 (https://doi.org/10.1109/ICSC58660.2023.10449868).

[19] M. Pouliquen *et al.*, "Identification of ARMAX Systems with Bounded Noise", *Journal of the Franklin Institute*, vol. 362, art. no. 108061, 2025 (https://doi.org/10.1016/j.jfranklin.2025.108061).

[20] H. El Maizi *et al.*, "Kernel-based Machine Learning Ellipsoidal Outer Bounding for Non-line-of-sight Outdoor and Indoor Channel Identification", *Telecommunication Systems*, vol. 89, art. no. 40, 2026 (https://doi.org/10.1007/s11235-026-01423-1).

[21] T.M. Lin, M. Nayeri, and J.R. Deller, "A Consistently Convergent OBE Algorithm with Automatic Estimation of Error Bounds", *International Journal of Adaptive Control and Signal Processing*, vol. 12, pp. 305–324, 1998 (https://doi.org/10.1002/(SICI)1099-1115(199806)12:4<305::AID-ACS491>3.0.CO;2-V).

[22] L. Pronzato and E. Walter, "Minimal Volume Ellipsoids", *International Journal of Adaptive Control and Signal Processing*, vol. 8, pp. 15–30, 1994 (https://doi.org/10.1002/acs.4480080103).

Hasna El Maizi, Ph.D.
LIMATI Laboratory
https://orcid.org/0009-0003-8047-7799
E-mail: hassnamaizi@gmail.com
University of Caen Normandy, Caen, France
https://www.normandie-univ.fr
Sultan Moulay Slimane University, Beni Mellal, Morocco
https://www.usms.ac.ma

Mathieu Pouliquen, Prof.
https://orcid.org/0000-0002-2186-986X
E-mail: mathieu.pouliquen@unicaen.fr
University of Caen Normandy, Caen, France
https://www.normandie-univ.fr

Rachid Fateh, Prof.
IMIS Research Laboratory
https://orcid.org/0000-0002-0574-2105
E-mail: fateh.smi@gmail.com
Sultan Moulay Slimane University, Beni Mellal, Morocco
https://www.usms.ac.ma
Ibn Zohr University, Agadir, Morocco
https://www.uiz.ac.ma

Miloud Frikel, Prof.
https://orcid.org/0000-0003-2178-1814
E-mail: miloud.frikel@ensicaen.fr
École Nationale Supérieure d'Ingénieurs de Caen, Caen, France
https://www.ensicaen.fr

Said Safi, Prof.
LIMATI Laboratory
https://orcid.org/0000-0003-3390-9037
E-mail: safi.said@gmail.com
Sultan Moulay Slimane University, Beni Mellal, Morocco
https://www.usms.ac.ma