HAL
open science

About frame identification of growth functions and robust prediction in bioprocess modeling

Emna Krichen, Alain Rapaport, E. Fouilland

To cite this version:

Emna Krichen, Alain Rapaport, E. Fouilland. About frame identification of growth functions and robust prediction in bioprocess modeling. 2019. hal-02095933v1

HAL Id: hal-02095933

https://hal.science/hal-02095933v1

Preprint submitted on 10 Apr 2019 (v1), last revised 3 Oct 2019 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

About frame identification of growth functions and robust prediction in bioprocess modeling

E. Krichen[1,2], A. Rapaport[1] and E. Fouilland[2]

[1] MISTEA, Univ. Montpellier, INRA, Montpellier SupAgro, France

e-mails: emna.krichen@supagro.fr, alain.rapaport@inra.fr

[2] MARBEC, Univ. Montpellier, CNRS, IRD, IFREMER, France

e-mail: eric.fouilland@cnrs.fr

April 10, 2019

Abstract

We address the problem of determining functional framing of experimental data points in view of robust time-varying predictions, which is of crucial importance in bioprocess monitoring. We propose a method that provides guaranteed functional bounds, instead of sets of parameters values of Monod or Haldane growth functions commonly used in bioprocess modeling. We illustrate the applicability of the method with bioreactor simulations in batch and continuous mode. We also present two extensions of the method adding flexibility in its application, and discuss its efficiency in providing guaranteed state estimations.

1 Introduction and motivations

Interval analysis has recently gained a lot of popularity for the robust estimation of state variables in dynamical models. Bioprocesses are typically subject to various kind of variations or disturbances which are not always easy to model [11, 17]. The idea to determine or compute bounds for uncertain systems is not new [38, 28, 48], but a recent renewed focus has been brought to estimate with intervals state variables or parameters from real-time observations, despite uncertainties on the dynamics [23, 24, 40, 26, 46]. In situations for which the usual random representations (based on probability measures, Gaussian processes, white noise...) are not appropriate or cannot be justified for modeling the perturbations, interval observers give a simple and attractive alternative [20]. Instead of (non-robust) single-valued observers or stochastic filters which both provide single estimates expected to represent average values (provided that hypotheses on the randomness are satisfied), interval observers give a set of plausible trajectories and/or parameters [20, 32, 36, 16, 51]. Typically, for a given deterministic model

$$\frac{dX}{dt} = F(t, X), \quad X \in \mathbb{R}^n, \tag{1}$$

where some terms of the map F are badly known or subject to fluctuations, and an observation vector

$$Y(t) = G(t, X(t)) \in \mathbb{R}^p \tag{2}$$

that is available at each time t, the interval approach consists in designing two estimators, a "lower" and a "upper" ones, of the form

$$\frac{dX^-}{dt} = F^-(t, X^-, Y(t)), \quad \frac{dX^+}{dt} = F^+(t, X^+, Y(t)), \tag{3}$$

when the dynamics F is monotone [50, 27], or in a coupled form

$$\frac{dX^-}{dt} = F^-(t, X^-, X^+, Y(t)), \quad \frac{dX^+}{dt} = F^+(t, X^-, X^+, Y(t)) \tag{4}$$

when F is non monotone (applying Muller's Theorem, see [38, 52, 32]). In this formulation, X^-, X^+ are vectors in $\mathbb{R}^n$ which provide a guaranteed framing of the solutions $X(\cdot)$ of (1) in the following sense

$$\begin{aligned} &\left\{X_i^-(0) \leq X_i(0) \leq X_i^+(0),\ (i = 1 \cdots n)\right\} \\ &\quad \Longrightarrow \left\{X_i^-(t) \leq X_i(t) \leq X_i^+(t),\ (i = 1 \cdots n)\right\}, \quad t > 0 \end{aligned} \tag{5}$$

This approach is particularly well suited to "slow" processes or dynamics with long transients, such as bioprocesses, because it guarantees bounds during the transients. It has been successfully applied to several models of bioprocesses [20, 13, 22, 42, 34, 41, 31, 18, 29, 14]. A possible drawback of this approach is that it could provide quite conservative bounds that could be then uninformative for practitioners. However, in recent years, much progress has been made to improve the width of the guaranteed intervals, playing with different structures of the systems [39, 30, 21, 53] (there exist also several results for the class of linear dynamics), changes of coordinates [44, 16], considerations of bundles of observers [12, 35] or with the help of purely numerical methods [25, 26, 40, 41, 32] based on interval analysis [37, 24].

In mass balance models for bioprocesses [11, 15], the most critical uncertainty is primarily found on the specific growth functions of the micro-organisms, often denoted by $\mu(\cdot)$. Consider for instance the classical chemostat model

$$\frac{dX}{dt} = F(t, X) := \begin{bmatrix} \mu(t, X)X_1 - D(t)X_1 \\ -\frac{1}{r}\mu(t, X)X_1 + D(t)(X_2^{in} - X_2) \end{bmatrix} \tag{6}$$

where X_1, X_2 denote the concentrations of biomass and substrate, respectively. The dilution rate $D(\cdot)$, the input substrate concentration X_2^{in} and the yield factor r, are supposed to be known. Several contributions among the ones cited previously have considered that the effective growth functions $\mu(\cdot)$ depends on time and on the substrate concentration $s = X_2$ in an unknown manner, but are bounded by two "extreme" functions $\mu^-(\cdot)$, $\mu^+(\cdot)$ so that the inequalities

$$\mu^-(s) \leq \mu(t, s) \leq \mu^+(s), \quad s \geq 0 \tag{7}$$

are fulfilled at time t. Robust state estimations can then be derived from the knowledge of $\mu^-(\cdot)$ and $\mu^+(\cdot)$ (instead of $\mu(\cdot)$), designing the maps F^-, F^+

in (3)-(4), as illustrated in [43, 10, 42]. However, one may wonder how to obtain these functional bounds from experimental data. Let us underline that it is particularly relevant to make a good choice of the functions μ^-, μ^+ as it impacts on the width of the frame (5). Take as an example the class of Monod growth functions

$$\mu(s) = \frac{\mu_{max} s}{s + K_s} \tag{8}$$

for which uncertainty is typically on the parameters μ_{max} and K_s. In the context of state or parameter estimation, several studies have focused on developing techniques to improve the computational efficiency when fitting models to measurement data, assuming arbitrary but bounded error distribution [8, 6, 3, 5, 7, 4]. Central of these approaches is overcoming convergence to local optima and inferring joint confidence intervals or regions [8, 9, 7]. It consists then in obtaining either parameter intervals $[\mu^-_{max}, \mu_{max}+]$, $[K_s^-, K_s^+]$ in $\mathbb{R}_+$ or a region of $\mathbb{R}_+ \times \mathbb{R}_+$. Differently from these techniques of guaranteed parameters estimations, the approach we propose here revolves around a functional framing. In fact, it is no longer a matter of looking for a set of parameters, but rather ensuring that there would be functions $\mu(\cdot)$ (depending on time in an unknown way) that guarantee that experimental data can be generated by a functional frame. This amounts to look for functional intervals $[\mu^-(\cdot), \mu^+(\cdot)]$ within a given class of functions. Although we shall consider classes of functions described by parameters, such as μ_{max} and K_s for the Monod functions, we shall simply look for two functions $\mu^-(\cdot)$, $\mu^+(\cdot)$ belonging to this class. Each of them is given by a parameters pair (μ^-_{max}, K_s^-), (μ^+_{max}, K_s^+), but this does not means that we shall consider that the unknown $\mu(\cdot)$ is a Monod function with unkwnown parameters $(\mu_{\max}, K_s)$ belonging to a particular subset. We rather state that the unknown growth function is *any* function $\mu(\cdot)$ verifying the condition (7). In particular $\mu(\cdot)$ is not necessarily a Monod function. Moreover, let us underline that the parameters are not necessarily such that $\mu^+_{max} > \mu^-_{max}$, $K_s^+ > K_s^-$ to obtain inequalities (7). To our knowledge, the question of determining functional intervals from experimental data in this way has not been yet addressed in the literature. Functional bounds rather than a parameters set appear to be better suited to the the method of interval estimators described above in the context of bioprocesses.

The purpose of the present work is to propose a generic method to fit two models (an "upper" and a "lower" one) that wrap data within lower and upper functions $\mu^-(\cdot)$ and $\mu^+(\cdot)$ instead of a single average $\mu(\cdot)$. To avoid too conservative bounds, we also introduce some flexibility in the application of the method for choosing the framing quality, either considering different upper and lower classes of growth functions, or forcing to have tightened bounds across the data sets, or both. We apply our framing method on synthetic data generated by random Monod [1] or Haldane [2] growth functions, this later one including the effect of substrate inhibition. Then, we compute time-varying predictions through numerical simulations of batch and continuous operating mode, providing bounded state estimations. As recalled previously, we mainly target situations for which the usual statistical hypotheses (reproducibility, uniform distribution of samples, unbiased noise...) are not necessarily met to justify the usual identification methods (least-squares, coefficient of determination, confidence intervals, maximum likelihood... see for instance [19, 49]). The approach

we propose is purely geometrical and does not rely on any statistical property, even though there are some similitude with least-square methods in the choice of the fitting criterion.

The paper is organized as follows. In Section 2, we present the method in a general framework and its theoretical justification (Proposition 1).In Section 3, we provide two extensions of the method while we show the application to growth curves in Section 4. Then, in Section 5, we report numerical data and simulations for two kinds of bioprocesses (batch and continuous) before we come to a conclusion and some perspectives.

2 Presentation of the method

We describe here the method in a general setting of an unknown scalar function $f : x \mapsto y = f(x)$, to be bounded by two functions f^-, f^+, given a set of n experimental measurements

$$\mathcal{S} := \{(x_i, y_i)\}_{i \in 1 \cdots n}.$$

We consider a class $\mathcal{C}$ of scalar functions parameterized by a vector of parameters $p \in P$, where P is a subset of $\mathbb{R}^m$:

$$\mathcal{C} := \bigcup_{p \in P} \{f(x, p) : x \in \mathbb{R} \mapsto \mathbb{R}\}. \tag{9}$$

and require minimal regularity conditions.

Assumption 1. *P is a compact subset of $\mathbb{R}^m$. f is continuous with respect to $(x, p) \in \mathbb{R} \times P$.*

Let us denote the range interval of x in the set $\mathcal{S}$ by

$$I := \left[\min_{i=1\cdots n} x_i \ , \ \max_{i=1\cdots n} x_i\right].$$

For each $p \in P$, we define the functions

$$L(p) = \sum_{i=1}^{n} \min(y_i - f(x_i, p), 0)^2, \quad U(p) = \sum_{i=1}^{n} \max(y_i - f(x_i, p), 0)^2 \ . \tag{10}$$

Note that $L(p) = 0$ amounts to say that the graph of the function $f(\cdot, p)$ is below $\mathcal{S}$. Similarly, $U(p) = 0$ amounts to say that the graph of the function $f(\cdot, p)$ is above $\mathcal{S}$. Therefore, any pair (p^-, p^+) such that

$$L(p^-) = U(p^+) = 0 \tag{11}$$

provides a frame of $\mathcal{S}$ in the sense that one has

$$f(x_i, p^-) \leq y_i \leq f(x_i, p^+), \quad i = 1 \cdots n. \tag{12}$$

Several pairs $(p^-, p^+) \in P^2$ could satisfy the frame condition (11), and provide frames that could be more or less loose. We look now for the tightest frames.

For each pair $(p^-, p^+) \in P^2$, we define a "distance" function:

$$D(p^-, p^+) := \sum_{i=1}^{n} \big(f(x_i, p^+) - f(x_i, p^-)\big)^2 \tag{13}$$

and for each $\gamma \geq 0$, we consider the criterion

$$J(p^-, p^+, \gamma) := L(p^-) + U(p^+) + \gamma D(p^-, p^+). \tag{14}$$

We then define the family of optimization problems

$$J^\star(\gamma) := \min_{(p^-, p^+) \in P^2} J(p^-, p^+, \gamma), \quad \gamma \geq 0 \tag{15}$$

parameterized by γ. Note that J is continuous w.r.t. (p^-, p^+) and P is a compact set, which guarantee that the minimum in $J^\star(\gamma)$ exists for any γ.

Let us also denote the classical least-square criterion by

$$J_{LS}(p) := \sum_{i=1}^{n} \big(y_i - f(x_i, p)\big)^2$$

and define the number

$$J^\star_{LS} := \min_{p \in P} J_{LS}(p).$$

We first assume that the class $\mathcal{C}$ is rich enough to wrap all the data.

Assumption 2. *There exists $(p^-, p^+) \in P^2$ such that* (12) *is fulfilled.*

We shall consider cases for which all the data do not belong exactly to the graph of single function in $\mathcal{C}$ (otherwise the classical least-square method would do the job of providing $\mu(\cdot) = \mu^-(\cdot) = \mu^+(\cdot)$).

Assumption 3. *There does not exist $p \in P$ such that one has $y_i = f(x_i, p)$ for any $i \in \{1, \cdots, n\}$.*

We give now our main result concerning the construction of the functional framing.

Proposition 1. *Under Assumptions 1-2-3, the supremum*

$$\sup \big\{\gamma \geq 0,\ \exists (p^-, p^+) \in P^2 \ \textit{s.t.}\ J^\star(\gamma) = \gamma D(p^-, p^+)\big\} \tag{16}$$

is reached for $\gamma^\star \in [0, \frac{1}{2}]$ and any pair (p^-, p^+) such that $J(p^-, p^+, \gamma^\star) = J^\star(\gamma^\star)$ verifies the framing (12)*, minimizing the function $D(p^-, p^+)$ among all pairs $(p^-, p^+) \in P^2$ satisfying the frame condition* (11)*.*

Proof. For convenience, define for any $\gamma \geq 0$ the set

$$E(\gamma) := \big\{(p^-, p^+) \in P^2 \text{ s.t. } J^\star(\gamma) = \gamma D(p^-, p^+)\big\}.$$

Remark first that having $p^- = p^+ = p$ gives $J(p, p, \gamma) = J_{LS}(p)$ whatever is γ, which is the usual least-squares criterion. Let $\tilde{p}$ be a solution of the least-square fitting, that is $\tilde{p}$ satisfying

$$J_{LS}(\tilde{p}) = J^\star_{LS} := \min_{p \in P} J_{LS}(p).$$

As $\mathcal{S}$ does not belong to the graph of a single function in $\mathcal{C}$ (Assumption 3), one has $J^\star_{LS} > 0$. For any pair (p^-, p^+) that satisfies the frame condition (11), one can write

$$\begin{aligned} J(p^-, p^+, \gamma) &= \gamma D(p^-, p^+) \\ &= \gamma \sum_{i=1}^{n} \Big(\big(f(x_i, p^+) - y_i\big) + \big(y_i - f(x_i, p^-)\big) \Big)^2 \\ &= \gamma \sum_{i=1}^{n} \big(y_i - f(x_i, p^+)\big)^2 + \big(y_i - f(x_i, p^-)\big)^2 \\ &\qquad\qquad +2\left(f(x_i, p^+) - y_i\right)\left(y_i - f(x_i, p^-)\right) \\ &= \gamma\Big(J_{LS}(p^-) + J_{LS}(p^+) + 2\left(f(x_i, p^+) - y_i\right)\left(y_i - f(x_i, p^-)\right) \Big) \end{aligned}$$

Property (11) implies that each term $(f(x_i, p^+) - y_i)\,(y_i - f(x_i, p^-))$ is non-negative, and one has then the inequality

$$J(p^-, p^+, \gamma) \geq \gamma(J_{LS}(p^-) + J_{LS}(p^+)) \geq 2\gamma J_{LS}(\tilde{p}) = 2\gamma J(\tilde{p}, \tilde{p}, \gamma)$$

which shows that for $\gamma > \frac{1}{2}$, an optimal solution for (15) cannot verify the frame condition (11). On another hand, for $\gamma = 0$, any pair (p^-, p^+) satisfying the frame condition (11) gives $J^\star(0) = 0$ and is thus optimal. We deduce that one has

$$\gamma^\star = \sup\{\gamma \geq 0,\ E(\gamma) \neq \emptyset\} \in \left[0, \tfrac{1}{2}\right]$$

Let us show now that this supremum is reached. Consider the subsets

$$P^+ = \bigcup_{i=1}^{n} \{p \in P,\ f(x_i, p) \geq y_i\}, \quad P^- = \bigcup_{i=1}^{n} \{p \in P,\ f(x_i, p) \leq y_i\}$$

By Assumption 2, P^- and P^+ are non-empty. Thanks to Assumption 1, P^- and P^+ are closed sets and the number

$$D^\star = \min\{D(p^-, p^+),\ (p^-, p^+) \in P^- \times P^+\}$$

is well defined. Moreover, there exist $(p^-_\star, p^+_\star) \in P^- \times P^+$ such that $D^\star = D(p^-_\star, p^+_\star)$, and by Assumption 3, $D^\star$ is positive.

Clearly, one has $E(\gamma) \subset P^- \times P^+$ for any $\gamma \geq 0$. If $J^\star(\gamma) < \gamma D^\star$, then for any $(p^-, p^+) \in P^- \times P^+$, one has $J(p^-, p^+, \gamma) = \gamma D(p^-, p^+) \geq \gamma D^\star > J^\star(\gamma)$ and $E(\gamma)$ is thus empty. If $J^\star(\gamma) \geq \gamma D^\star$, then the pair $(p^-_\star, p^+_\star)$ realizes the equality $J^\star = J(\gamma, p^-_\star, p^+_\star) = \gamma D(p^-_\star, p^+_\star)$ and one has $(p^-_\star, p^+_\star) \in E(\gamma)$ with $J^\star(\gamma) = \gamma D^\star$. Therefore, one has

$$\{\gamma \in [0, \tfrac{1}{2}]\ E(\gamma) \neq \emptyset\} = \{\gamma \geq 0,\ J^\star(\gamma) = \gamma D^\star\}$$

and as $J^\star$ is continuous, we deduce that $\max\{\gamma \in [0, \frac{1}{2}],\ E(\gamma) \neq \emptyset\}$ exists.

Finally, the pair $(p^-_\star, p^+_\star)$ satisfies (12) and minimizes $D(\cdot)$ among $P^- \times P^+$, which ends the proof. $\square$

In practice, one may easily approximate $\gamma^\star$ with a good accuracy by a dichotomy method, as follows.

Algorithm 1 Estimate $\gamma^\star$

Require: $n > 1$

$\gamma^- \leftarrow 0$, $\gamma^+ \leftarrow 1/2$
for $i = 1 \cdots n$ **do**
  $\gamma \leftarrow \frac{\gamma^- + \gamma^+}{2}$
  let p^-, p^+ be such that $J(p^-, p^+, \gamma) = J^\star(\gamma)$
  if $D(p^-, p^+) > 0$ **then**
    $\gamma_+ \leftarrow \gamma$
  else
    $\gamma_- \leftarrow \gamma$
  end if
end for
return γ^-, γ^+

This algorithm gives numbers γ^-, γ^+ such that $\gamma^\star \in [\gamma^-, \gamma^+]$ with $\gamma^+ - \gamma^- < 2^{-n-1}$.

For each pair (p^-, p^+) verifying the frame condition (11), we measure the width of the frame by the area delimied by the graphs of the functions $f(\cdot, p^+)$ and $f(\cdot, p^-)$:

$$A(p^-, p^+) := \int_{x \in I} f(x, p^+) - f(x, p^-)\, dx \tag{17}$$

This allows to discriminate between between pairs $(p^-_\star, p^+_\star)$ realizing the minimum of $D(\cdot)$ when they are non unique. Moreover, one can associate to the class of functions $\mathcal{C}$ a "quality" as the number

$$Q(\mathcal{C}) := \min \left\{ A(p^-, p^+);\ (p^-, p^+) \in P^2 \text{ s.t. } J(p^-, p^+, \gamma^\star) = J^\star(\gamma^\star) \right\}.$$

This allows to compare the adequacy of each class. For two classes $\mathcal{C}_1$, $\mathcal{C}_2$ of functions, we shall say that $\mathcal{C}_1$ gives a better framing of the set $\mathcal{S}$ than the class $\mathcal{C}_2$ when one has $Q(\mathcal{C}_1) < Q(\mathcal{C}_2)$.

Finally, let us draw some comparisons with the classical least-square method.

Lemma 1. *Under Assumptions* (1)-(2)-(3)*, one has*

$$\lim_{\gamma \to +\infty} J^\star(\gamma) = J^\star_{LS}\ . \tag{18}$$

Proof. Notice that $\gamma \mapsto J^\star(\gamma)$ is non decreasing, as the family of functions $(J(\cdot, \cdot, \gamma))$ is non-decreasing w.r.t. γ, and is bounded from above by $J^\star_{LQ}$. Therefore the limit of $J^\star(\gamma)$ when γ tends to $+\infty$ exists.

Consider a sequence of optimal pairs $(P^{\star -}_k, P^{\star +}_k)$ for $\gamma = k$. As P is compact, the sequence converges when $k \to +\infty$, up to a sub-sequence, to a certain $(P^{\star -}_\infty, P^{\star +}_\infty) \in P^2$ such that $J(P^{\star -}_\infty, P^{\star +}_\infty, \infty) = \lim_{\gamma \to +\infty} J^\star(\gamma)$. This implies that one has necessarily $D(P^{\star -}_\infty, P^{\star +}_\infty) = 0$, that is $f(x_i, P^{\star -}_\infty) = f(x_i, P^{\star +}_\infty)$ for any $i = 1 \cdots n$. Consequently, one has $J(P^{\star -}_\infty, P^{\star +}_\infty, \infty) \geq J^\star_{LS}$ and we obtain that equality (18) is verified. □

Consequently, our family of optimization problems generalizes the well-known least-square optimization for large values of γ.

It may happen that the data set $\mathcal{S}$ is too poor or that the class $\mathcal{C}$ is over-parameterized. We shall say that $\mathcal{S}$ is *rich enough* for $\mathcal{C}$ when the following property is fulfilled.

$$\left\{f(x_i, p^a) = f(x_i, p^b);\ i = 1 \cdots n\right\} \Rightarrow p^a = p^b\ . \tag{19}$$

Under this last condition, any converging sequence of optimal pairs $(P_k^{\star -}, P_k^{\star +})$ satisfies $P_\infty^{\star -} = P_\infty^{\star +}$, i.e. the optimal frames converge when $\gamma \to +\infty$ to single curves, solutions of the least-square optimization.

3 Extensions of the method

The extensions we present here bring some additional flexibility to the method, in terms of choice of the class of functions and width of frames. For simplicity of the presentation, we have not incorporated them in the presentation of the overall method. Both extensions are illustrated in Section 5.

3.1 Framing with different upper and lower classes

There is no a priori reason to impose that upper and lower functions in the frame (12) are sought among the same class of functions. Therefore, we can consider without loss of generality two classes

$$\mathcal{C}^- := \bigcup_{p \in P^-} \{f^-(x,p) : x \in \mathbb{R} \mapsto \mathbb{R}\}, \quad \mathcal{C}^+ := \bigcup_{p \in P^-} \{f^+(x,p) : x \in \mathbb{R} \mapsto \mathbb{R}\}$$

instead of the single class $\mathcal{C}$ considered in (9). Here, P^- and P^+ are two subsets of $\mathbb{R}^{m^-}$ and $\mathbb{R}^{m^+}$, such that Assumption 1 is fulfilled for each pair (P^-, f^-), (P^+, f^+). The dimension m^- is not necessarily equal to m^+. In Section 5, we shall see that this allows to consider less parameters altogether.

In short, we do not give here the corresponding assumptions and statement, as its simply consists in replacing $P \times P$ by $P^- \times P^+$ in Assumption 2 and Proposition 1 of Section 2. We shall also consider that Assumption 3 is verified for both P^- and P^+.

3.2 Relaxing with one or more points out of the frame

In practice, the frame provided by Proposition 1 might be too conservative as it wraps all the experimental data. Practitioners might want to have tightened bounds, allowing to have one or several points unwrapped. They could of course choose themselves which point(s) to be redrawn and relaunch the method on a smaller data set. We show here how to extend in a simple way the method to have this done in a systematic way, i.e. as an automatic method choosing which data point(s) can be forgotten to obtain tightened bounds, depending on the maximal number of points that one accepts to be unwrapped.

For each pair (p^-, p^+) in P^2 (or in $P^- \times P^+$), we define the sub-set of indices

$$O(p^-, p^+) = \left\{i \in \{1, \cdots, n\},\ y_i \notin [f(x_i, p^-), f(x_i, p^+)\right\}$$

As a generalization of the optimization problem considered in Section 2, we consider the following family of problems

$$\gamma_j^\star := \sup\left\{\gamma \geq 0,\ \exists (p^-, p^+) \in P^2 \text{ s.t. } J^\star(\gamma) = J(p^-, p^+, \gamma),\ \#O(p^-, p^+) \leq j\right\}$$

for $j \in \{0, n-1\}$, where $\#E$ denotes the cardinality of a set E.

Clearly, one has $\gamma_j^\star > \gamma_0^\star = \gamma^\star$ for any $j > 0$, where $\gamma^\star$ is given by Proposition 1. Then, for each $\gamma_j^\star$, there exits a pair (p_j^-, p_j^+) solution of the problem (15) with $\gamma = \gamma_j^\star$, which provides a frame with at most j unwrapped points. However, notice that the optimization problem with $j > 0$ do not guarantee that the pair (p_j^-, p_j^+) provides the minimal value of $D(p^-, p^+)$ among the pairs that exclude exactly j points, differently to the original method for $j = 0$ (see Proposition 1). For this later problem, one has to consider a different criterion which consists in computing the functions L, U are D only for indices out of the subset $O(p^-, p^+)$. This last subset do depend on the pair (p^-, p^+), which makes the computation of the optimization problem much heavier. Instead, we propose here a simpler relaxation approach which simply consists in changing the condition on line 5 in Algorithm 2 by

$$\#O(p^-, p^+) > j.$$

The family (p_j^-, p_j^+) provides then a sensitivity of the framing with respect to the number of unwrapped points. This will be illustrated in Section 5.

4 Application to the framing of growth curves

The knowledge of the growth functions is very useful in microbiology to predict the evolution of micro-organisms and the performances of the bioprocesses, with models such as (6):

$$\left\{\begin{array}{rcl} \dot{x} & = & \mu(s)x - Dx \\ \dot{s} & = & -\frac{1}{r}\mu(s)x + D(s_{in} - s) \end{array}\right. \tag{20}$$

where x and s stand here for the biomass and substrate concentrations, respectively. Typically, experiments to determine the unknown growth function $\mu(\cdot)$ are drawn as follows.

- either in batch culture i.e. with $D = 0$. Experiments consist in choosing different values s_i of initial substrate concentration $s(0) = s_i$ and determining the quantity $\mu_i = \frac{d}{dt}\ln(x(t))|_{x=0}$, measuring the time variation of $x(\cdot)$. Data points (s_i, m_i) are then expected to belong to the graph of the unknown function $\mu(\cdot)$.

- either in continuous culture i.e. with $D > 0$ (assumed to be constant or slowly varying). Experiments consist in choosing different values D_i of D and wait the system to be at a quasi-steady state $(x^\star, s^\star)$. Then, the data point $(s_i, \mu_i) = (s^\star, D_i)$ is expected to belong to the graph of the unknown function $\mu(\cdot)$.

With the set of data points (s_i, μ_i), one usually conducts a non-linear regression to adjust the parameters of a growth model with least-square criterion [47], such

as for instance the Monod growth function (8), or the Haldane one

$$\mu(s) = \frac{\bar{\mu} s}{K + s + s^2/K_i}. \tag{21}$$

describing the effect of growth inhibition for large concentrations of substrate. However, it has been underlined that the usual R-squared determination coefficient is not appropriate to measure the quality of the adjustment in this non-linearity context [45]. More importantly, reproducibility and environmental fluctuations (such as pH, temperature, light...) from one experiment to another one are common issues in microbiology.

Here, we shall consider that the experiments have been sufficiently numerous to have faced the various possible kinds of fluctuations, so that future uses of the micro-organisms might be subject to similar fluctuations or combinations of them at different times. This hypothesis is important to justify the framing approach, that we illustrate on synthetic data in the next section.

Let us now show how one can use the lower and upper functions $\mu^-(\cdot)$, $\mu^+(\cdot)$ to obtained guaranteed predictions of $x(\cdot)$ and $s(\cdot)$. Assume that yield coefficient r, dilution rate D (possibly time-varying) and initial condition (x_0, s_0) of model (20) are known. The, one can easily show that the variable $z(t) = x(t) + rs(t)$ is solution of $\dot{z} = D(rs_{in} - z)$. Therefore, one has

$$z(t) = rs_{in} + (x_0 + rs_0 - rs_{in})e^{-Dt}, \quad t \geq 0. \tag{22}$$

We distinguish now the predictions of $x(t)$ and $s(t)$.

1. Prediction of $x(t)$. From $s(t) = \frac{1}{r}(z(t) - x(t))$, one obtains that $x(\cdot)$ is solution of the (non-autonomous) scalar differential equation:

$$\dot{x} = \mu\left(\tfrac{1}{r}(z(t) - x)\right) x - Dx$$

Then, following [52, Th. VIII, p. 95], a frame $x^-(t) \leq x(t) \leq x^+(t)$ is obtained for any $t \geq 0$ when $x^-(\cdot)$, $x^+(\cdot)$ are solutions of

$$\begin{cases} \dot{x}^- = \mu^-\left(\frac{1}{r}(z(t) - x^-)\right) x^- - Dx^-, \ x^-(0) = x_0 \\ \dot{x}^+ = \mu^+\left(\max\left(\frac{1}{r}(z(t) - x^+), 0\right)\right) x^+ - Dx^+, \ x^+(0) = x_0 \end{cases} \tag{23}$$

2. Prediction of $s(t)$. Similarly, one has $x(t) = z(t) - rs(t)$ and $s(\cdot)$ solution of the scalar dynamics

$$\dot{s} = -\tfrac{1}{r}\mu(s)(z(t) - rs(t)) + D(s_{in} - s)$$

which gives a frame $s^-(t) \leq s(t) \leq s^+(t)$ where $s^-(\cdot)$, $s^+(\cdot)$ are solutions of

$$\begin{cases} \dot{s}^- = -\frac{1}{r}\mu^+(s^-)(z(t) - rs^-) + D(s_{in} - s^-), \ s^-(0) = s_0 \\ \dot{s}^+ = -\frac{1}{r}\mu^-(s^+)\max\left(z(t) - rs^+, 0\right) + D(s_{in} - s^+), \ s^-(0) = s_0 \end{cases} \tag{24}$$

Remark that uncertainties on the initial condition $(x_0, s_0) \in [x_0^-, x_0^+] \times [s_0^-, s_0^+]$ can be also incorporated in the framing, considering

$$z^-(t) = rs_{in} + (x_0^- + rs_0^- - rs_{in})e^{-Dt}, \ z^+(t) = rs_{in} + (x_0^+ + rs_0^+ - rs_{in})e^{-Dt}$$

instead of $z(t)$ in systems (23)-(24) and initializations $x^-(0) = x_0^-$, $x^+(0) = x_0^+$, $s^-(0) = s_0^-$, $s^+(0) = s_0^+$.

5 Numerical illustrations

In the section we illustrate the framing method on synthetic data and its use for predictions as follows.

1. We generate a random set $\mathcal{P}$ of parameters of the growth models Monod or Andrews, from which we generate a set $\mathcal{S}$ of "synthetic" data points (s_i, μ_i), corrupted by some noise. The set $\mathcal{S}$ is considered to represent a plausible set of experimental data.

2. We apply our framing method to the data set $\mathcal{S}$ for determining the best lower and upper functions $\mu^-(\cdot)$, $\mu^+(\cdot)$ (cf Proposition 1).

3. We use the functions $\mu^-(\cdot)$, $\mu^+(\cdot)$ to compute time varying intervals $[x^-(\cdot), x^+(\cdot)]$, $[s^-(\cdot), s^+(\cdot)]$ for the concentrations $x(\cdot)$, $s(\cdot)$, as solutions of equations (22)-(23)-(24).

4. We simulate the dynamical model (20) where the function $\mu(\cdot)$ changes randomly with time among functions (8) or (21) with parameters within $\mathcal{P}$, and compare the solution $x(\cdot)$, $s(\cdot)$ during the transients with the lower and upper estimates provided by the previous step.

We have also determined the average growth function and its corresponding trajectory to appreciate its position within the intervals. The optimization problem (15) has been solved numerically with the `fmincon` function of `Matlab` software.

5.1 Example of a batch process

We have generated randomly ten growth curves (see Fig. 1, left) and one hundred data corrupted with noise (see Fig. 1, right).

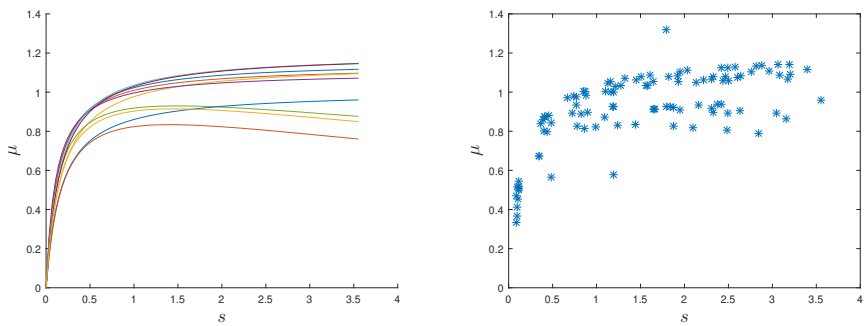

Figure 1: Random growths (left) and random data points (s_i, μ_i) (right).

We have wrapped all the data points among the class of Monod functions, using the method exposed in Section 2, which gives the functions μ^-, μ^+ whose graphas are depicted on Fig. 2-left. In addition, the graph of the growth curve $\bar{\mu}$ determined by the usual least-square method is plotted in dashed line. We have then simulated a trajectory of the model (20) in batch mode (i.e. for $D = 0$) with a random growth $\mu(\cdot)$, along with the solutions of the frame dynamics

(23)-(24) (see Fig. 2-right). The value of the yield coefficient r has been chosen equal to 0.1. One can notice that the solution provided by the average growth $\bar{\mu}$ (in dashed line) is not centered in the frame. This is due to the non-linearity of the dynamics. This shows the interest of the information provided by the guaranteed interval approach, confidence intervals of least-square method being always centered about an average trajectory.

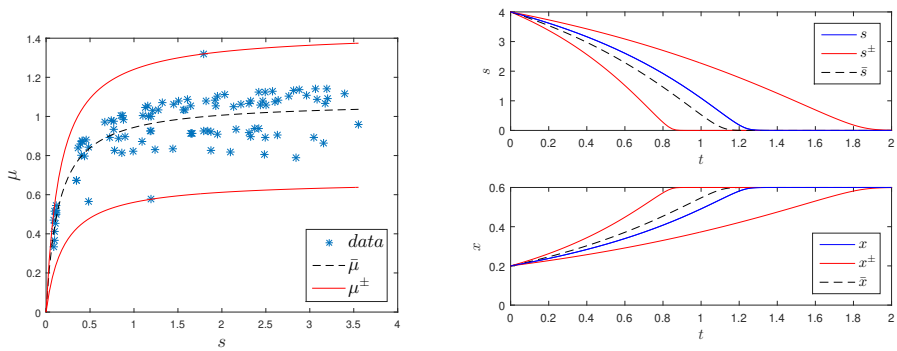

Figure 2: Wrapping of the data (left) and dynamical simulation (right).

One may want to relax the frame considering that some data points are most probably aberrant. As explained in Section 3, the extension of the method chooses which points to be out of the frame. On Fig 3 one can appreciate the impacts on the interval width, allowing only two data points to be unwrapped.

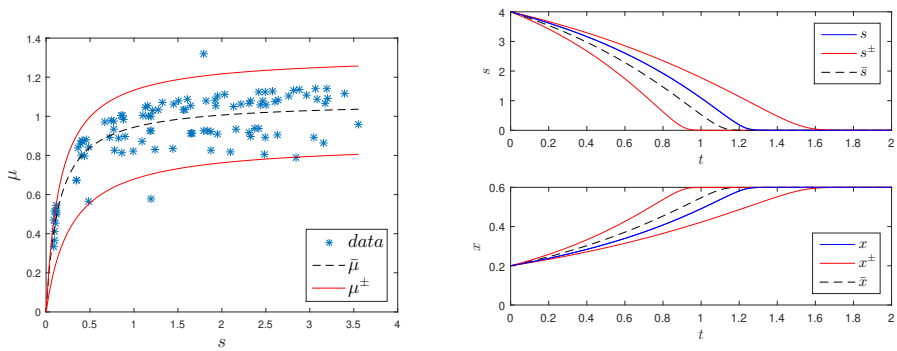

Figure 3: Data wrapping (left) and corresponding simulation (right) allowing two data points to be out of the frame.

In batch process, the asymptotic value of the biomass x is known, equal to $x_\infty = x_0 + rs_0$ according to (22), but guaranteed intervals allow to estimate the time necessary to reach a given level of production $x_l < x_\infty$, which is of primer utility in industrial applications.

5.2 Example of a continuous process

We have generated another set of random growth curves and random data (see Fig. 4).

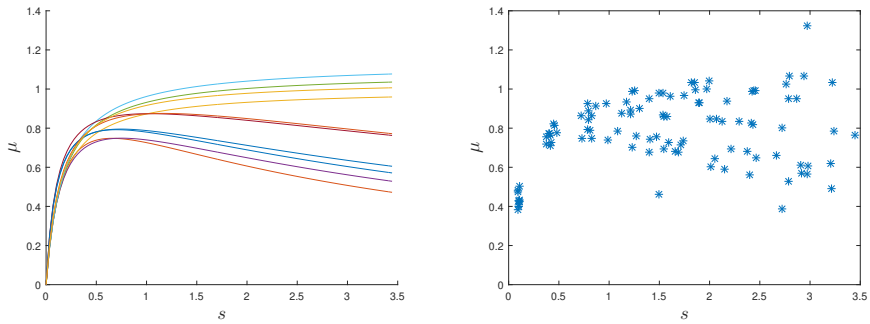

Figure 4: Random growth curves (left) and random points (s_i, μ_i) (right).

As previously, we have wrapped the data points in between the graphs of two functions μ^-, μ^+ among the class of Monod growth functions, depicted on Fig. 5-left (the average growth determined by the least-square criterion is plotted in dashed line). We have then considered the model (20) in continuous mode for the input parameters $D = 0.7$ and $s_{in} = 4$ (with the same yield coefficient $r = 0.1$ than before). Fig. 5-right shows a trajectory for a random growth $\mu(\cdot)$ lying inside the guaranteed time-varying intervals computed with the functions μ^-, μ^+.

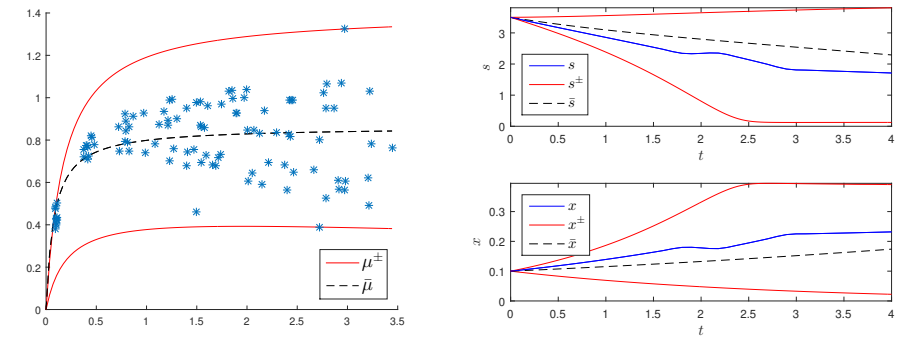

Figure 5: Wrapping with Monod growth functions (left) and associated dynamical simulation (right)

To illustrate the flexibility of the extensions of the method presented in Section 3, we have looked for the upper function μ^- in the class of Andrews functions (21), allowing thus non monotonic growth curves (see Fig. 6). The area criterion (17) gives an improvement of the fitting (2.66 instead of 3.01). However, this does not impact too much the frame estimation as this concerns only small values of s. In addition, we have relaxed the framing allowing at most three points to be unwrapped (see Fig. 7), which has reduced more significantly

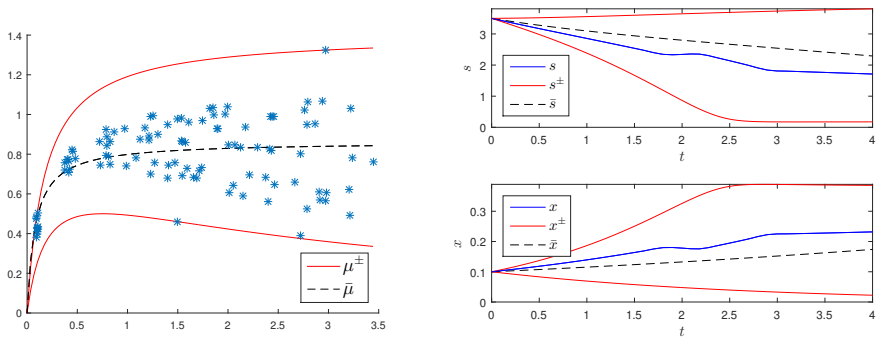

Figure 6: Wrapping with Monod and Andrews growth functions (left) and associated simulation (right)

the frame estimation. Here also, one can see that the interval method is relevant

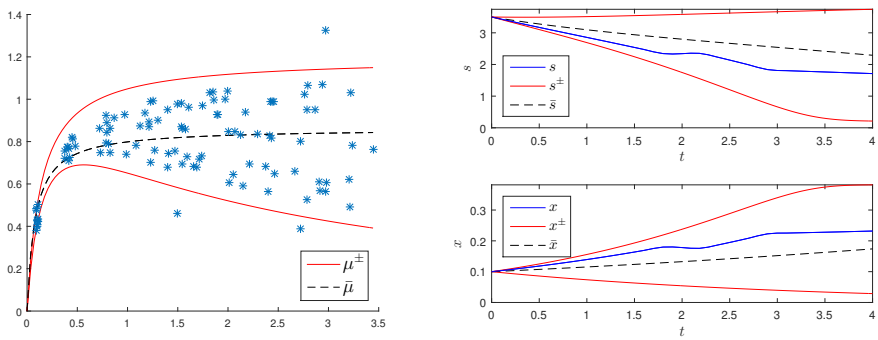

Figure 7: Data wrapping (left) and corresponding simulation (right) requiring three data points to be out of the frame.

to obtain safe transients estimates, while the use of least-square method is more questionable in the context of unpredictable variations.

6 Conclusion and perspectives

We have proposed a methodology for the construction of functional frames from data points and applied it microbial growth functions, which are often the major source of variability in microbiology. However, these growth functions may depend on other variables, such as biomass concentration in the Contois model, or could be multi-valued, such as in the Droop model where the nutrient uptake function $\rho(\cdot)$ has to be identified concomitantly to the growth rate function $\mu(\cdot)$. Extensions of the method in these multi-valued contexts might be the matter of future works.

Acknowledgments

This work was supported by the LabEx NUMEV incorporated into the I-Site MUSE, and the ADEME French Agency.

References

[1] Monod, Jacques *La technique de culture continue, théorie et applications*, Ann. Inst. Pasteur, 79, 390–410,1950.

[2] Andrews, John F *A mathematical model for the continuous culture of microorganisms utilizing inhibitory substrates.* Biotechnology and Bioengineering, 10, 6, 707–723, 1968.

[3] Jaulin, Luc and Walter, Eric *Guaranteed nonlinear parameter estimation from bounded-error data via interval analysis*, Mathematics and computers in simulation, 35, 2, 123–137, 1993.

[4] Marvel, Skylar W and Williams, Cranos M *Set membership experimental design for biological systems*, BMC systems biology, 6, 1, 21, 2012.

[5] Milanese, Mario *Properties of least squares estimates in set membership identification*, 31, 2, 327–332, 1995.

[6] Moore, Ramon *Parameter sets for bounded-error data*, 34, 2, 113–119, 1992.

[7] Perić, Nikola D and Paulen, Radoslav and Villanueva, Mario E and Chachuat, Benoît *Set-membership nonlinear regression approach to parameter estimation*, 70, 80–95, 2018.

[8] Walter, Eric and Piet-Lahanier, Hélène *Estimation of parameter bounds from bounded-error data: a survey*, 32, 5-6, 449–468, 1990.

[9] Mukkula, Anwesh Reddy Gottu and Paulen, Radoslav *Model-based design of optimal experiments for nonlinear systems in the context of guaranteed parameter estimation*, Computers & Chemical Engineering, 99, 198–213, 2017.

[10] Alcaraz-Gonzalez, V., Harmand, J., Rapaport, A., Steyer J.P., Gonzalez-Alcarez, V. and Pelayo-Ortiz, C. *Software sensors for higly uncertain WWTPs: a new approach based on intervals observers*, Water Research, 36, 2515–2524, 2002.

[11] Bastin, G. and Dochain, D. *On-line estimation and adaptive control of bioreactors*, Elsevier, 1990.

[12] Bernard, O. and Gouzé, J.-L. *Closed loop observers bundle for uncertain biotechnological models.* Journal of process control 14 (7), 765–774, 2004.

[13] Bernard, O., Gouzé, J.-L. and Hadj-Sadok, Z. and *Observers for the biotechnological processes with unknown kinetics. Application to wastewater treatment.* In: Proceedings of the 39th IEEE Conference on Decision and Control, Sydney (Australia), 2000.

[14] Bunciu, E. *Biomass Estimation for an Anaerobic Bioprocess Using Interval Observer* In: Proceedings of the 3rd International on Intelligent Decision Technologies (IDT'2011) Watada J., Phillips-Wren, G., Jain, L. and Howlett, R. (Eds), Smart Innovation, Systems and Technologies, Springer, 95–102, 2011.

[15] Dochain, D. *Automatic Control of Bioprocesses.* (Editor) Control Systems, Robotics and Manufacturing series, ISTE-Wiley, 2008.

[16] Efimov, D., Raissi, T., Chebotarev, S., Zolghadri, A. *Interval state observer for nonlinear time varying systems.* Automatica, 49(1), 200–205, 2013

[17] Goffaux, G. and Vande Wouwer, A. *Bioprocess State Estimation: Some Classical and Less Classical Approaches.* In: Control and Observer Design for Nonlinear Finite and Infinite Dimensional Systems. Springer, 111–128, 2005.

[18] Goffaux, G., Vande Wouwer, A. and Bernard, O. *Continuous-discrete interval observers for monitoring microalgae cultures.* Biotechnology progress 25 (3), 667–675, 2009.

[19] Goodwin, G. and Payne, R. *Dynamic System Identification, Experiment Design and Data Analysis.* Academic Press, New York, 1977.

[20] Gouzé, J.-L., Rapaport, A. and Hadj-Sadok, Z. *Interval observers for uncertain biological systems*, Ecological modelling, 133, 45–56, 2000.

[21] Gucik-Derigny, D., Raissi, T; and Zolghadri, A. *A note on interval observer design for unknown input estimation.* Int. J. Control, 89 (1), 25–37, 2016.

[22] Hadj-Sadok, Z. and Gouzé J.-L. *Estimation of uncertain models of activated sludge processes with interval observers.* Journal of Process Control, 11 (3), 299–310., 2001.

[23] Jaulin L. *Nonlinear Bounded-Error State Estimation of Continuous-Time Systems.* Automatica, 38 (6), 1079–1082., 2002.

[24] Jaulin, L., Kieffer, M., Didrit, O. and Walter, E. *Applied Interval Analysis: with examples in parameter and state estimation, robust control and robotics.* Springer-Verlag, 2001.

[25] Kieffer, M. and Walter, E. *Guaranteed nonlinear state estimator for cooperative systems.* Numerical Algorithms, 37, 187–198, 2004.

[26] Kieffer, M., Walter, E. and Simeonov, I. *Guaranteed non-linear parameter estimation for continuous-time dynamical models.* In: Proceedings 14th IFAC Symposium on System Identification, Newcastle (Australia), 843–848, 2006.

[27] De Leenheer, P., Angeli D. and Sontag, E., *A Tutorial on Monotone Systems - With an Application to Chemical Reaction Networks.* In Proc. 16th Int. Symp. Macd thematical Theory of Networks and Systems (MTNS), Leuven (Belgium), 2004.

[28] Lohner, R. *Computation of guaranteed enclosures for the solutions of ordinary initial and boundary value-problem.* In: Cash, J. and Gladwell I. (Eds.) Computational Ordinary Differential Equations, Clarendon Press, Oxford, 425–435, 1992.

[29] Mairet, F., Moisan, M. and Bernard, O. *Interval observer-based estimator of specific growth rate in bioreactors.* Journal Européen des Systèmes Automatisés (JESA), Lavoisier, 44 (4-5), 493–507, 2010.

[30] Mazenc, F. and Bernard, O. *ISS interval observers for nonlinear systems transformed into triangular systems.* Int. J. Robust Nonlinear Control, 24 (7), 1241–1261, 2014.

[31] Meslem, N., Ramdani, N. and Candau, Y. *Interval Observers for Uncertain Nonlinear Systems. Application to Bioreactors.* In: Proceedings of the 17th World Congress, The International Federation of Automatic Control Seoul (Korea) July 6–11, 9667–9672, 2008.

[32] Meslem, N., Ramdani, N. and Candau, Y. *Approximation garantie de l'espace d'état atteignable des systémes dynamiques continus incertains. Encadrement par intervalle de flot d'état.* Journal Euroéen des Systémes Automatisés (JESA), 43, 589–614, 2009.

[33] Meslem, N., Ramdani, N. and Candau, Y. *Using hybrid automata for set-membership state estimation with uncertain nonlinear continuous-time systems.* Journal of Process Control, 20, 481–489., 2010.

[34] Moisan, M. and Bernard, O. *Interval observers for non monotone systems. Application to bioprocess models.* IFAC Proceedings 38 (1), 43–48, 2005.

[35] Moisan, M. and Bernard, O. *Near optimal interval observers bundle for uncertain bioreactors.* Automatica, 45 (1):291–295, 2009

[36] Moisan, M. and Bernard, O. *Robust interval observers for global Lipschitz uncertain chaotic systems.* Systems & Control Letters 59 (11), 687-694, 2010.

[37] Moore, R. *Interval analysis.* Prentice-Hall, Englewood Cliffs, 1966.

[38] Muller, M. *Uber das Fundamentaltheorem in der Theorie der gewohnlichen Differential-gleichungen.* Math. Z., 619–645, 1927.

[39] Raissi, T., Efimov, D. and Zolghadri, A. *Interval state estimation for a class of nonlinear systems.* IEEE Transactions on Automatic Control, 57 (1), pp. 260–265, 2012.

[40] Raissi, T., Ramdani, N. and Candau, Y. *Set membership state and parameter estimation for systems described by nonlinear differential equations.* Automatica, 40 (10), 1771–1777., 2004.

[41] Raissi, T., Ramdani, N. and Candau, Y. *Bounded-error moving horizon state estimator for non-linear continuous-time systems : application to a bioprocess system.* Journal of Process Control, 15 (5), 537–545, 2005.

[42] Rapaport, A and Dochain, D. *Interval observers for biochemical processes with uncertain kinetics and inputs.* Mathematical biosciences, 193 (2):235–253, 2005.

[43] Rapaport, A. and Harmand, J. *Robust regulation of a class of partially observed nonlinear continuous bioreactors.* Journal of Process Control, 12, 291–302, 2002.

[44] Rapaport, A. and Gouzé, J.L. *Parallelotopic and practical observers for nonlinear uncertain systems.* Int. Journal. Control, 76 (3), 237–251, 2003.

[45] Ratkowsky, D. *Model fitting and uncertainty*, In: Modeling Microbial Responses in Food. McKellar, R., Lu, X. (Eds.), CRC Press, Boca Raton, 152–196, 2004.

[46] Rauh, A. and Auer , E. (Eds) *Modeling, Design, and Simulation of Systems with Uncertainties.* Mathematical Engineering series, Springer 2011.

[47] Robinson., J. *Determining microbial kinetic parameters using non-linear regression analysis.* Adv. Microb. Ecol., 8, 61–114 (1985).

[48] Singer, A. and Barton, P. *Bounding the solutions of parameter dependent nonlinear ordinary differential equations.* SIAM J. Sci. Comput., 27 (6), 2167–2182, 2006.

[49] Schittkowski, K. *Numerical Data Fitting in Dynamical Systems, A Practical Introduction with Applications and Software*, Springer, 2002.

[50] Smith, H. *Monotone Dynamical Systems.* American Mathematical Society, Providence, 1995.

[51] Thabet, R., Raissi, T., Combastel, C., Efimov, D. and Zolghadri, A. *An effective method to interval observer design for time-varying systems.* Automatica, 50 (10), 2677–2684, 2014.

[52] Walter, W. *Differential inequalities and maximum principles: theory, new methods and applications.* Nonlinear analysis, Theory, Methods and Applications, 30(8), 4695–4711, 1997.

[53] Zheng, G., Efimov, D. and Perruquetti, W. *Design of interval observer for a class of uncertain unobservable nonlinear systems.* Automatica, 63, 167–174, 2016.