Nonlinear Robust Identification using Evolutionary Algorithms. Application to a Biomedical Process

J.M. Herrero; X. Blasco; M. Martínez; C. Ramos; J. Sanchis

Department of Systems Engineering and Control, Polytechnic University of

Valencia, Camino de vera s/n, 46022 Valencia, Spain;

Tel.: +34-96-3879571, Fax: +34-96-3879579

e-mail of corresponding author: juaherdu@isa.upv.es

Abstract

This work describes a new methodology for Robust Identification (RI), meaning the identification of the parameters of a model and the characterization of uncertainties. The alternative proposed handles non-linear models and can take into account the different properties demanded by the model. The indicator that leads the identification process is the identification error (IE), that is, the difference between experimental data and model response. In particular, the methodology obtains the feasible parameter set (FPS set of parameters values which satisfy a bounded IE) and a nominal model in a non-linear identification problem. To impose different properties on the model, several norms of the IE are used and bounded simultaneously. This improves the model quality, but increases the problem complexity. The methodology proposes that the Robust Identification problem is transformed into a multimodal optimization problem with an infinite number of global minima which constitute the FPS. For the optimization task, a special Genetic Algorithm ($\epsilon-$GA), inspired by

Multiobjective Evolutionary Algorithms, is presented. This algorithm characterizes the FPS by means of a discrete set of models well distributed along the FPS. Finally, an application for a biomedical model which shows the blockage that a given drug produces on the ionic currents of a cardiac cell is presented to illustrate the methodology.

Key words: Robust Identification, Multimodal Optimization, Multiobjective Optimization, Evolutionary Algorithms, Biomedical Processes.

1 Introduction

Obtaining mathematical models which describe systems or process behaviour is a fundamental task in many scientific areas, especially in biomedicine where the process is the human being. When the model is obtained from first principles, the problem finally consists of identifying the model parameters through process information which can be obtained from experimental information regarding the process inputs and outputs.

The fact that the process behaviour is not completely known and that available data is insufficient, or unreliable, forces the identified parameters to have uncertainty, which should be taken into account when the model is used for predictions, controller design, or other tasks. The work of identifying the nominal model and its uncertainty is called robust identification (RI).

Two different approaches are possible in RI: stochastic or deterministic. In the former, the identification error (IE), meaning the difference between the process output measurements and the model simulated outputs, is assumed to be modelled as a random variable with several statistical properties. Under

this approach, it is possible to use classical techniques of identification (Walter and Pronzalo, 1997; Ljung, 1999) to obtain the nominal model and its uncertainty; which is related to the covariance matrix of the estimated parameters. When these assumptions do not work, the deterministic approach can be more appropriate (Norton, 1987; Walter and Piet-Lahanier, 1990; Milanese and Vicino, 1991), where the identification error, although unknown, is assumed to be bounded.

The objective of the deterministic approach is to obtain the nominal model and its uncertainty; or directly, the feasible parameter set (FPS), i.e. the parameter set which keeps the IE bounded for certain IE functions or norms, and their bounds.

When the model has linear parameters, the FPS is, if it exists, a convex polytope. This polytope may be complex because the number of vertices can grow exponentially as the number of observations increases, and so the complexity involved in obtaining the polytope can be considerable. The polytope is often approximated using orthotopes (Belforte et al., 1990), ellipsoids (Fogel and Huang, 1882), or parallelotopics (Chisci et al., 1998); and these generally result in a more conservative characterization of the FPS.

When the model is non-linear, the FPS may be a non-convex, and even disjoint and polytope - and this makes it more difficult to find a tight characterization of the FPS. Some techniques: such as interval computation (Walter and Kieffer, 2003); support vector machine (Keesman and Stappers, 2004), and others, can be used. However, these techniques suffer limitations (the type of function for bounding the IE, the inability to characterize a non-convex or disjoint FPS), or their use is complicated when the model is complex

(non-differentiable with respect to its parameters, discontinuities in parameters and/or signals, etc.).

To overcome these handicaps, a more flexible and general methodology for characterizing FPS is presented. It can identify many processes and characterize convex, non-convex, and even disjoint FPS. In addition, several norms can be taken into account at the same time. This enables, for instance, bounding the IE for each experimental sample and its integral simultaneously; as well as the consideration of independent norms for each output. The practical sense of simultaneous norms is justified: for example, it would be useful if the model predictions attempt to satisfy a limited maximal error (∞-norm) and - at the same time - find a good average fitting between model and experiment (absolute norm).

The proposed methodology is based on the optimization of a function which is built from IE norms and bounds, and whose global minima will characterize the FPS. It will be a multimodal function, which can be non-convex and/or present local minima, and therefore classical optimizers (for instance, SQP [1]) can be inappropriate.

The FPS depends on the norms used to bound the IE, and especially on their corresponding bounds [2].

To select the bounds, a priori process knowledge (for instance, non-modelled

_____

[1] Sequential Quadratic Programming is a variant of the iterative Gauss-Newton optimization method.

[2] For instance, the FPS obtained by taking into account the ∞-norm and the η bound corresponds to the parameter set which forces the IE not to be greater than η in all the output samples.

dynamics) and noise characteristics must be used. However, as this can be a hard task, the bound is often selected by taking into account the desired performance for the model predictions. Low values for the bounds could achieve an empty FPS, whereas high values could provide a more conservative FPS; so the IE bound selection is a critical decision.

For selecting the bounds and avoiding an $FPS = \emptyset$, a procedure which uses solutions (Pareto Front information) of a multiobjective problem associated with norms used will be proposed. The Pareto Front is obtained by the simultaneous minimization of the IE norms, through a multiobjective optimization (MO).

In relation to the nominal model, a well-known estimate is the Chebyshev centre (Garulli et al., 2000) of the FPS. This is the best worst case nominal model, in the sense that it minimizes the maximum distance to FPS. Obtaining the Chebyshev centre is sensitive to the bounds and besides, when the FPS is unavailable, it can become a difficult task, so other possibilities could be used: for instance; analytic centre (Bai, 1999); interpolatory projection; central projection; or restricted projection (Garulli et al., 2000).

In this work, a method to obtain a nominal model using a restricted interpolatory projection is presented. This nominal model is optimal from the point of view of both the IE and the estimation error in the parameter space.

The work is organized as follows: in section 2 a new global optimization technique for multimodal problems, which is required for solving the proposed RI methodology, is presented. The section describes the fundamentals of the algorithm developed (ϵ-GA). The proposed RI methodology is shown in section 3. Section 4 shows an example of modelling and RI of a process, in particu-

lar, the behaviour of a certain drug on the blockage of the ionic currents of a cardiac cell. The main conclusions are given in section 5.

2 ϵ-GA evolutionary algorithm

Before the detailed description of the algorithm it is important to define some related concepts and properties to satisfy. ϵ-GA is an evolutionary algorithm designed to optimize multimodal mono-objective functions which have an infinite number of global optima.

2.1 Concepts related to the ϵ-GA

The solution of the optimization problem consists of:

Definition 1 *(Global minimum set) Given a finite domain $D \subseteq \mathcal{R}^L$, $D \neq \emptyset$ and a function to optimize $J : D \rightarrow \mathcal{R}$, the set Θ^* will be the global minimum set of J, if and only if, Θ^* contains all the global optima of J.*

$$\Theta^* := \{\theta \in D : J(\theta) = J^*\}, \tag{1}$$

being J^* the global minimum of J for the searching space D.

From this definition, Θ^* is assumed to be a unique set and the best that can be achieved is to obtain a discretized approximation to Θ^* in the solution space D, that means, a finite set Θ^*_ϵ.

An important property to satisfy is the requirement that the algorithm obtains a well distributed solution (the points of the solution set have to cover, as

6

uniformly as possible, the space of the global minimum). Notice that the complete solution is not usually reachable (it is an infinite set) and the objective is to obtain a good approximation.

To achieve this, the solution space is divided by a grid into boxes of width ϵ_i for each dimension $i \in [1 \dots L]$ and the algorithm is forced to produce just one solution for each box. The solutions Θ^*_ϵ are forced to be well distributed and characterize Θ^*.

A practical approach to definitions required for the algorithm description is as follows (formal definition can be consulted in Herrero (2006)):

- Quasi-global minimum: a point with a value of J near to the global minimum. The distance to the global minimum value is set with the parameter δ.

$$\theta \text{ is a quasi-global minimum } \Leftrightarrow J(\theta) \leq J^* + \delta, \tag{2}$$

- Quasi-global minimum set: the set of quasi-global minima.
- Box: number of the zone in which the parameter space is divided by the grid. The size of the grid is set, for each dimension i, by the parameter ϵ_i.
- Box-representative: because several points can be at the same box, a method to characterize all of them is defined, the box-representative is a point of a box that characterizes all the other points of the box. The representative has the lowest value of J inside the box, and if several points have the same value the nearest to the geometrical center of the box is preferred.

    Notice that the box is defined as having a finite good discretization of the solution, and it is not necessary to have several points of the same box in the resulting solution.

- ϵ-global minimum set: a set of quasi-global minimum points which are box-representative. This set is not unique.

All these definitions help to describe the methodology for producing a solution (set of point Θ_ϵ^* that represents the solution Θ^*). In particular, the methodology on how to include points in Θ_ϵ^* can be set as follows: *"A point is included in Θ_ϵ^* if it is a quasi-global minimum and box-representative"*. In other words, the point is near enough to the solution and is well distributed (according to the defined boxes). Moreover, once a point is added to Θ_ϵ^* the whole set must be revised to remove those that do not satisfy the conditions any more. This means *"Points non-quasi global minimum or non box-representative are removed from Θ_ϵ^*"*.

Notice that for quasi-global minimum definition, it is necessary to know the global minimum J^*, but this value is usually unknown, instead J_Θ^{min} is used, the approximation whose value of the function J is the smallest among the set of evaluated points (Θ).

$$J_\Theta^{min} = \min_{\theta \in \Theta} J(\theta). \tag{3}$$

It can be proven (see Herrero (2006)) that with this inclusion procedure the contents of Θ_ϵ^* converge towards an ϵ-global minimum set; as long as $J_{\Theta_\epsilon^*}^{min}$ converges towards the global minimum J^*.

Finally, the effect of parameters ϵ_i and δ is described. Parameters ϵ_i show the desired discretization degree to apply to Θ_ϵ^* and they are directly related to the physical meaning of the parameters which define the searching space dimensions.

The parameter δ plays two roles related to convergence and diversity:

- A value $\delta \simeq 0$ improves the convergence and $\Theta_\epsilon^* \Rightarrow \Theta^*$, but worsens the approximation of Θ^* and so its characterization.

- On the contrary, a too high value of δ could cause the quasi-global minimum solutions of Θ_ϵ^* to distort Θ^* instead of characterizing it.

A good procedure to choose δ could consist of starting from a value $\delta = \delta_{ini}$ and modifying it during algorithm execution - (for instance, by using a decreasing exponential function) towards a value $\delta = \delta_{fin}$ low enough to make the quasi-global minimum solutions be near the global minimum solutions.

2.2 ϵ-GA description

The objective of the ϵ-Genetic Algorithm (ϵ-GA) is to provide an ϵ-global minimum set, Θ_ϵ^*. This algorithm is inspired in evolutionary optimization algorithms and shares the well known basic concepts of population, genetic operator, codification, etc.

ϵ-GA uses the populations $P(t)$, $A(t)$ and $G(t)$ (see figure 1):

(1) $P(t)$ is the main population and explores the searching space D. The population size is $Nind_P$.

(2) $A(t)$ is the archive where Θ_ϵ^* is stored. Its size $Nind_A$ is variable but bounded.

(3) $G(t)$ is an auxiliary population which is used to store the new individuals generated at each iteration by the algorithm. The population size is $Nind_G$.

The pseudocode of the ϵ-GA algorithm is given by:

```
1. t:=0

2. A(t):=∅

3. P(t):=ini_random(D)

4. eval(P(t))

5. A(t):=store(P(t),A(t))

6. mode:=exploration

7. while t<t_max do

8.     G(t):=create(P(t),A(t))

9.     eval(G(t))

10.    A(t+1):=store(G(t),A(t))

11.    P(t+1):=update(G(t),P(t))

12.    mode:=determinemode(P(t))

13.    t:=t+1

14. end while
```
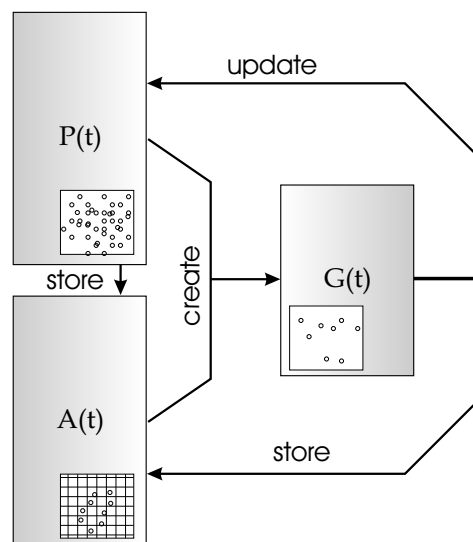

Figure 1. ϵ-GA algorithm structure.

The main steps of the above algorithm are detailed below:

Step 3. Population $P(0)$ is initialized with $Nind_P$ individuals, created inside the searching space D.

Steps 4 and 9. Function *eval* calculates the value of the fitness function $J(\theta)$ for every individual θ from $P(t)$ (step 4) or $G(t)$ (step 9).

Step 12. The function *determinemode* selects the algorithm operation mode between the exploration and exploitation modes. These modes affect how new individuals are created (function create). When the population $P(t)$ has converged, the exploitation mode must be selected, by using the difference between the best value

$$J_{P(t)}^{min} = \min_{\theta \in P(t)} J(\theta) \tag{4}$$

and the worst value

$$J_{P(t)}^{max} = \max_{\theta \in P(t)} J(\theta) \tag{5}$$

at iteration t. If $J_{P(t)}^{max} - J_{P(t)}^{min} < \delta$ the *exploitation* mode [3] will be selected, on the contrary, the *exploration* mode will be selected.

Step 5 and 10. Function *store* analyzes whether every individual of $P(t)$ (step 5) or $G(t)$ (step 10) must be included in archive $A(t)$. So the individual will have to satisfy the inclusion condition described and, according to this, other individuals will be removed. When including a new individual, if there is already a box-representative in the same box, then the nearest to the center

---

[3] If $J_{P(t)}^{min} = J^*$ all the individuals in $P(t)$ will be quasi-global minimum solutions.

of the box is preferred. So, a better distribution of the solutions inside the archive is achieved.

Step 8. Function *create* creates new individuals and stores them in population $G(t)$ using the following procedure:

(1) Two individuals are randomly selected, θ^{p1} from $P(t)$, and θ^{p2} from $A(t)$.

(2) If the algorithm operates in *exploration* mode, θ^{p2} is not altered, whereas if the mode is *exploitation*, the individual is mutated, according to:

$$\theta_i^{p2} = \theta_i^{p2} + N(0, \beta_{ini}). \tag{6}$$

(3) A random number $u \in [0 \ldots 1]$ is selected. If $u > P_{c/m}$ (crossover-mutation probability) step 4 (crossover) is taken, otherwise step 5 (mutation).

(4) θ^{p1} and θ^{p2} are crossed over by the extended linear recombination technique and two new individuals θ^{h1} and θ^{h2} are created [4] :

$$\theta_i^{h1} = \alpha_i(t) \cdot \theta_i^{p1} + (1 - \alpha_i(t)) \cdot \theta_i^{p2}, \tag{7}$$
$$\theta_i^{h2} = (1 - \alpha_i(t)) \cdot \theta_i^{p1} + \alpha_i(t) \cdot \theta_i^{p2}. \tag{8}$$

(5) θ^{p1} and θ^{p2} are mutated by random mutation with gaussian distribution [5] .

$$\theta_i^{h1} = \theta_i^{p1} + N(0, \beta1_i(t)), \tag{9}$$
$$\theta_i^{h2} = \theta_i^{p2} + N(0, \beta2_i(t)). \tag{10}$$

This procedure is repeated $Nind_G/2$ times until $G(t)$ is full.

---

[4] $\alpha_i(t)$ is a random value with uniform distribution $\in [-d(t), 1 + d(t)]$ and $d(t)$ is a parameter tuned by a decreasing exponential function $d(t) = \dfrac{d_{ini}}{\sqrt{1 + \left(\left(\frac{d_{ini}}{d_{fin}}\right)^2 - 1\right)\frac{t}{(tmax-1)}}}$.

[5] Variances $\beta1_i(t)$ and $\beta2_i(t)$ are expressed in percentage of $(\theta_{i\ max} - \theta_{i\ min})$ and are tuned by a function similar to the one used for tuning $d(t)$.

Step 10. Function *update* updates $P(t)$ with individuals from $G(t)$. One individual θ^G from $G(t)$ will be inserted in $P(t)$ and it will replace θ^p, $J(\theta^G) < J(\theta^p)$ being

$$\theta^p = \arg \max_{\theta \in P(t)} J(\theta) \tag{11}$$

so, the contents of $P(t)$ are converging.

Finally, when $t = t_{max}$, the individuals included in the archive $A(t)$ will be the solution Θ_ϵ^* to the multimodal optimization problem, being Θ the set of individuals generated by steps 3 and 8, that is,

$$\Theta = P(0) \bigcup \left(\bigcup_{0 \leq \tau < t_{max}-1} G(t) \right) \tag{12}$$

and being $\Theta \cap \Theta^* \neq \emptyset$.

3 Robust Identification problem

The task of identifying the nominal model and its uncertainty is called robust identification (RI). For problem statement, a model structure is assumed and the parameters have to be identified. The uncertainty is characterized with a set of possible value of the parameters called *Feasible Parameters Set* (*FPS*).

Assuming the following model structure:

$$\hat{\mathbf{y}}(t, \theta) = f(t, \mathbf{u}(t), \theta) \tag{13}$$

where:

- $f(.)$ is the model function.

- $\theta \in D \subset R^L$ is the vector [6] of the unknown model parameters.

- $\mathbf{u}(t) \in R^m$ is the vector of model inputs.

- $\hat{\mathbf{y}}(t, \theta) \in R^l$ is the vector of model outputs.

The objective is that the model behaviour (obtained by experiments) will be as similar as possible to the real process behaviour (obtained by simulation).

This objective can be achieved by a minimization of a function which penalizes the Identification Error (IE) for process outputs.

The identification error $\mathbf{e}_j(\theta)$ for the output $j \in [1 \dots l]$ is stated:

$$\mathbf{e}_j(\theta) = \mathbf{y}_j - \hat{\mathbf{y}}_j(\theta), \tag{14}$$

where:

- $\mathbf{y}_j = [y_j(t_1), y_j(t_2) \dots y_j(t_N)]$ are the process output j measurements [7] when the inputs $\mathbf{U} = [\mathbf{u}(t_1), \mathbf{u}(t_2) \dots \mathbf{u}(t_N)]$ are applied to the model.

- $\hat{\mathbf{y}}_j(\theta) = [\hat{y}_j(t_1, \theta), \hat{y}_j(t_2, \theta) \dots \hat{y}_j(t_N, \theta)]$ are the simulated model output j when the same inputs $\mathbf{U}$ are applied to the model [8].

To introduce desirable characteristics in the model it is very helpful to be able to bound IE in different ways. For instance, a practical and intuitive approach is to bound average IE and maximum IE simultaneously, meaning that on average, the model will fit experimental data and the maximum error is limited. Then the identification error must be bounded by several norms [9]

---

[6] θ, $\mathbf{x}(t)$, $\mathbf{u}(t)$ and $\hat{\mathbf{y}}(t, \theta)$ are all column vectors.

[7] $\mathbf{y}(t) \in \mathcal{R}^l$ is the column vector of process outputs.

[8] N is the measurements number of each output and input. The interval between measurements is constant $t_i = i \cdot T_s$, being T_s the sample time.

[9] In a more general case, it would be possible to use bounds on any function.

simultaneously.

Let N denote a p-norm of the identification error vector for an output j as:

$$N(\theta) = \|\mathbf{e}_j(\theta)\|_p, \tag{15}$$

If s norms must be bound simultaneously, the feasible parameter set FPS_i is consistent with a specific norm N_i and bound η_i for $i \in A := [1, 2, \ldots s]$, is defined as:

$$FPS_i := \{\theta \in D : N_i(\theta) \le \eta_i, \eta_i > 0\}. \tag{16}$$

The FPS_i is the set of points in the search space that verifies the constraint established with the bounded norm.

To characterize FPS, it is important to define its boundary:

$$\partial FPS_i := \{\theta \in D : N_i(\theta) = \eta_i, \eta_i > 0\}. \tag{17}$$

Therefore, the FPS for all the norms simultaneously is stated as:

$$FPS := \{\bigcap_{i \in A} FPS_i\} = \{\theta \in D : \forall i \in A, N_i(\theta) \le \eta_i, \eta_i > 0\}. \tag{18}$$

and its boundary

$$\partial FPS := \{\theta \in D : \exists i | N_i(\theta) = \eta_i \wedge N_j(\theta) \le \eta_j\} \tag{19}$$

with $\eta_i, \eta_j > 0$ and $i, j \in A$.

Then the RI problem solution is FPS, and in particular ∂FPS. The method proposed to solve it is to reformulate the RI problem as a multimodal opti-

mization problem, and to solve it using the $\epsilon - GA$ described in last section to solve it.

3.1 RI problem as a multimodal optimization problem

To characterize the FPS, and in particular its boundary ∂FPS, a function $J(\theta)$ is stated in such a way that its global minima constitutes the ∂FPS and the FPS constitutes quasi-global minimum solutions, the parameter δ is the one used by $\epsilon - GA$ to define quasi-global minimum.

$$J(\theta) := \begin{cases} \sum_B J_i & if \ B(\theta) \neq \emptyset \\ \\ \min(\delta, \prod_A J_i) \ if \ B(\theta) = \emptyset \end{cases} \tag{20}$$

where:

$$B(\theta) := \{i \in A : N_i(\theta) > \eta_i\}, \tag{21}$$
$$J_i(\theta) = |N_i(\theta) - \eta_i|. \tag{22}$$

Some of the properties of function $J(\theta)$ are:

(1) Global minimum of $J(\theta)$ is $J^* = 0$ and marks the contour of FPS (∂FPS). Notice that for the points of the FPS ($B(\theta) = \emptyset$) the value of $J(\theta)$ is $\min(\delta, \prod_A J_i)$ and exactly for points of ∂FPS one, or more, of J_i are $J_i(\theta) = 0$.

(2) $J(\theta) < \delta$ when $\theta \in FPS$, therefore, it is ensured that these solutions are quasi-global minimum ones and they will be never removed from archive $A(t)$ by algorithm ϵ-GA. Besides, they will not prevail over the solutions $\theta \in \partial FPS$ either, therefore the boundary characterization will

16

be a priority.

Below, a case with two norms is evaluated to clarify $J(\theta)$ building. To simplify visualization, parameter space dimension is $L = 1$ (that is, $\theta \in \mathcal{R}$). η_1, η_2, $J_1(\theta)$ and $J_2(\theta)$ are set as:

$$\eta_1 = 20 \Rightarrow J_1(\theta) = |N_1(\theta) - 20|, \tag{23}$$
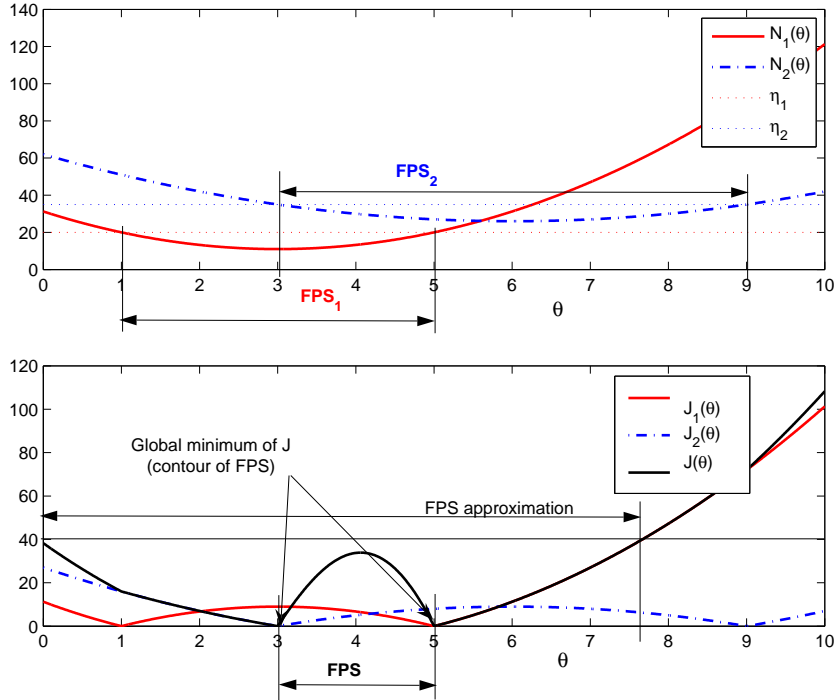$$\eta_2 = 35 \Rightarrow J_2(\theta) = |N_2(\theta) - 35|. \tag{24}$$

Figure 2. $J(\theta)$ building for a case with 2 norms to minimize simultaneously. $\delta = 40$.

Figure 2 shows J building for the case of $\delta = 40$. Notice that the global minimum of J are ∂FPS independently of the value of δ. The parameter δ is set to characterize the complete FPS and all values of J under δ constitute a FPS approximation. If δ is too high, the approximation is bad (see figure 2).

Results of the example are:

$$FPS_1 := \{\theta : \theta \in [1 \dots 5]\}, \tag{25}$$

$$FPS_2 := \{\theta : \theta \in [3 \dots 9]\}, \tag{26}$$

and then,

$$FPS := \{\theta : \theta \in [3 \dots 5]\}, \tag{27}$$
$$\partial FPS = \{3, 5\}, \quad J(3) = J(5) = 0. \tag{28}$$

For this particular case, where $L = 1$, ∂FPS is a finite set with only two points.
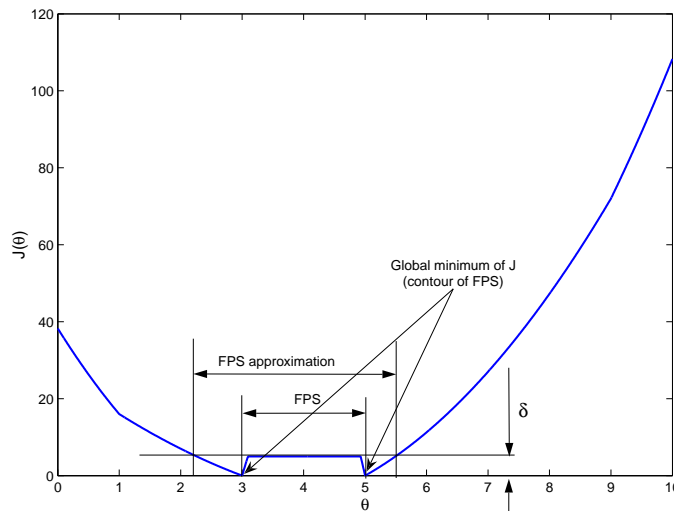
Figure 3. $J(\theta)$ building for a case with 2 norms to minimize simultaneously. $\delta = 5$.

Adjusting $\delta = 5$ gives a better approximation of the entire FPS, see figure 3. ∂FPS is always characterized by global minimum of J.

3.2 Bound selection

To select the bounds η_i on N_i, a priori process knowledge (for instance, non-modelled dynamics) and noise characteristics must be taken into account. However, this can be a difficult task, and the bounds may often be selected according to the desired performance for the model predictions. An inappropriate selection of the bounds might result in a conservative FPS if too high

18

values are chosen, or an $FPS = \emptyset$ if the values are too low.

According to other authors (Walter and Piet-Lahanier, 1991), to avoid this last case ($FPS = \emptyset$) when a unique norm $N_1(\theta)$ is used, it is useful to select the minimization bound by the $N_1(\theta)$, that is, the lower bound $\eta_1^{min} = \min_\theta N_1(\theta)$ and an $FPS \neq \emptyset$ is satisfied if $\eta_1 \geq \eta_1^{min}$.

When several norms are simultaneously taken into account, the fact of selecting $\eta_i \geq \eta_i^{min}$ (being $\eta_i^{min} = \min_\theta N_i(\theta)$) does not imply that $FPS \neq \emptyset$.

This work proposes an alternative method to select the η_i bounds by the simultaneous optimization of the N_i norms, through the following multiobjective optimization problem:

$$\min_{\theta \in D} \mathbf{J}(\theta); \quad \mathbf{J}(\theta) = \{N_1(\theta), N_2(\theta), \dots, N_s(\theta)\}. \tag{29}$$

The optimization problem solution is the Pareto optimum solutions set $\hat{\mathbf{\Theta}}_P$ (or a discrete approximation $\hat{\mathbf{\Theta}}_P^*$). Once the optimization problem is solved (applying a multiobjective optimization algorithm, for instance, ϵ-MOGA (Herrero et al., 2005)), it is possible to use the Pareto Front information $\mathbf{J}(\hat{\mathbf{\Theta}}_P^*)$ for selecting the η_i bounds - as is shown below. Figure 4 shows the case in which two norms, N_1 and N_2 of the identification error, are used. If bounds are selected as η_1 and η_2 the FPS is not empty, points of FPS belong to shadowed area in figure 4. However, a selection of bound as $\bar{\eta}_1$ and $\bar{\eta}_2$, even if they satisfy $\bar{\eta}_1 > \eta_1^{min}$ and $\bar{\eta}_2 > \eta_2^{min}$, produces an empty FPS.
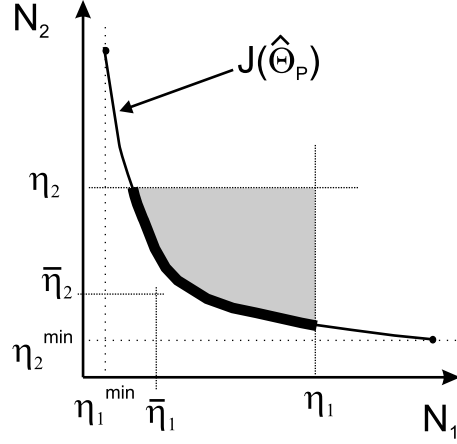
19

Figure 4. Example of bound selection to avoid empty FPS. $\{N_1(FPS), N_2(FPS)\} \in$ shadowed area.

3.3   Nominal model selection

Once characterized the FPS through the FPS^* (discretized approximation obtained with ϵ-GA), it is possible to approximate the worst case optimum nominal model by calculating the Chebyshev centre of the FPS^* as:

$$\hat{\theta}_c^* = arg \min_{\theta \in \mathbf{D}} \max_{\overline{\theta} \in \mathbf{FPS^*}} \|\theta - \overline{\theta}\|_2, \tag{30}$$

This nominal model selection is based exclusively on geometrical layout of the FPS set. The option proposed in this work includes information about optimality in the IE. Because the multiobjective problem has been solved to select bounds, the Pareto set is available. The proposal is to choose the nearest point to the Chebyshev center belonging to Pareto set ($\hat{\Theta}_P^*$) and FPS^*.

$$\hat{\theta}_{pi}^* = \min_{\theta \in \hat{\Theta}_P^* \bigcap FPS} \|\theta - \hat{\theta}_c^*\|. \tag{31}$$

The advantage of this nominal selection is that θ_{pi}^* belongs to the Pareto set, and so is an optimal solution with respect to IE.

20

3.4   FPS validation

A good practice in model identification requires a model validation using different experimental data.

Calling FPS_{ide} to the feasible parameter set determined via robust identification, using the experimental data $\Omega_{ide} = \{\mathbf{Y}_{ide}, \mathbf{U}_{ide}\}$, the s norms N_i and their bounds η_i. One method of validation consists of checking whether the FPS_{ide} contains models which are consistent with new data $\Omega_{val} = \{\mathbf{Y}_{val}, \mathbf{U}_{val}\}$. This means that the FPS obtained by process identification with data $\Omega = \{\Omega_{ide}, \Omega_{val}\}$ would be $FPS \neq \emptyset$.
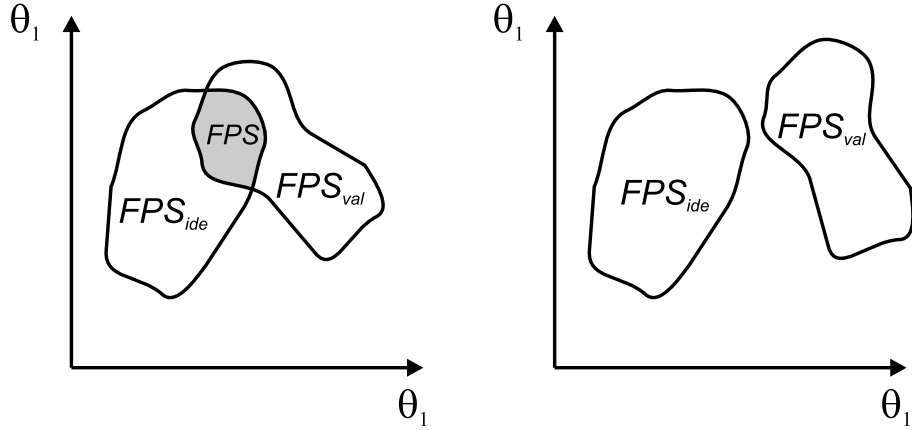
Figure 5. Validation process. On the left, the FPS_{ide} is validated, since $FPS \neq \emptyset$. On the right, the FPS_{ide} is invalidated since $FPS = \emptyset$.

In figure 5 two cases are shown. In the first example, there are models in the FPS_{ide} which also belong to the FPS_{val} (set consistent with Ω_{val} and with the same s norms N_i and bounds η_i used for FPS_{ide}), and therefore, the FPS_{ide} is validated; and in the second example, this does not occur and so the FPS_{ide} is invalidated [10].

_____

[10] In the same way, it is possible to validate a certain model $\hat{\theta} \in FPS$ by checking whether $\hat{\theta} \in FPS_{val}$.

If the FPS_{ide} is validated, the final FPS will be $FPS = FPS_{ide} \bigcap FPS_{val}$. It is not necessary to obtain the FPS_{val}, but only to maintain in the FPS those models from FPS_{ide} which are consistent with the data Ω_{val}. Since the finite set FPS_{ide}^* is available, obtaining the FPS^* is easy, because it is only necessary to simulate the models $\theta \in FPS_{ide}^*$ (using Ω_{val}) and choose those which satisfy $N_i(\theta) \leq \eta_i \ \forall i \in A$.

If the FPS_{ide} is invalidated, several options could be considered:

- To increase some, or all, the η_i bounds until the FPS_{ide} can be validated with data Ω_{val}.
- To modify the model (for instance, by adding part of the non-modelled dynamics) until the FPS_{ide} is validated.

In this second option, it is not necessary to increase the η_i bounds and so model prediction performance does not deteriorate as occurs with the first action. However, the model would surely be more complex.

3.5   Summary

The proposed methodology to solve an RI problem follows several steps:

(1) Collect experimental data for identification and validation.
(2) Establish a model structure and the parameters to identify.
(3) Select desirable properties for the model, that is, select the norms to consider.
(4) Choose adequate bounds using the results of the multiobjective problem solution as proposed. If resulting bounds are above the desired bounds,

22

go to step (2). This case occurs when desired bounds are too restrictive, or the structure model is not adequate for the process.

(5) Obtain an FPS with the algorithm $\epsilon - GA$.

(6) Validate FPS with other experimental data and if necessary recalculate FPS.

(7) Obtain, if required, a nominal model.

4 RI of a biomedical model

4.1 Biomedical model

The objective of this section is to show the equations which model the periodic iteration between a certain drug and a certain ion channel in cardiac cells. The drug can plug the channel and modify its action potential (AP) [11], helping to correct certain pathologies which occur in normal cardiac behaviour [12].

There are many alternatives (see Cardona (2005) and its references) to model the interaction between the drug and the receptor (ion channel), although in this work the Guarded Receptor (GRT) hypothesis is adopted. The channel configuration depends on the AP and involves accesible channels (A) which can be bound (B) by the drug, and inaccesible channels (I) (due to the channel gates which impede the blockage).

---

[11] The AP is the potential difference between the intracellular and the extracellular media.

[12] This work is part of a collaboration project with the Dept. of Electronic Eng. of the UPV.
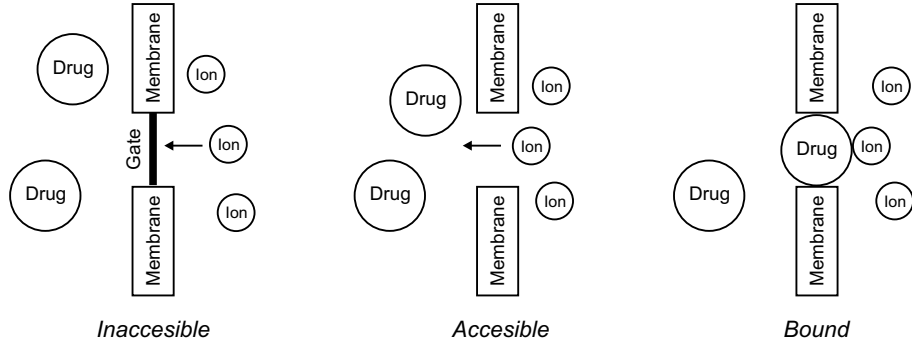
Figure 6. GRT model of interaction between the drug and the channels, and the gates which impede access.

The interaction dynamics between the I and A states is rapid in comparison to those between A and B. If the cell is stimulated by a pulse, the channel behaviour can be described as a two-step process, involving an activation state (activation interval, t_a) and a recovery channel state (recovery interval, t_r)

$$U + D \underset{L_i}{\overset{K_i}{\Longleftrightarrow}} B, \tag{32}$$

where i represents the state (activation a or recovery r), and U the unbound channels. K_i and L_i are the apparent rates of binding and unbinding between the drug and the channel, which are used to compare the drugs - Starmer (1988).

The time course of the fraction of bound channels is adjusted to the following first order system.

$$\frac{db}{dt} = K_i D(1 - b) - L_i b. \tag{33}$$

Figure 7 shows the exponential evolution (typical of first order systems) of

24

$b(t)$ depending on the interval (activation or recovery).
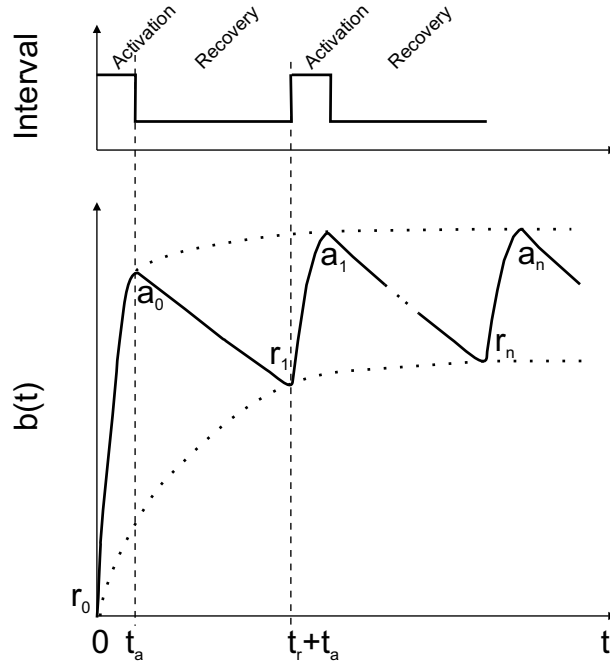
Figure 7. Evolution of $b(t)$ depending on the interval (activation or recovery). a_n and r_n represent the fraction of bound channels just prior to each time interval. The sequences (a_n, r_n) also follow an exponential pattern.

The solution to equation (33) is:

$$b(t) = b_{i,\infty} + (b_0 - b_{i,\infty})e^{-\lambda_i t}, \tag{34}$$

where $b_{i,\infty} = \frac{K_i D}{K_i D + L_i}$ is the fraction of bound channels at $t = \infty$, $\lambda_i = K_i D + L_i$ is the time constant and b_0 is the fraction of bound channels at $t = 0$.

The fraction of bound channels just prior to each time interval can be described by a sequence of recurrence equations such as:

$$r_n = a_{n-1}e^{-\lambda_r t_r} + r_\infty(1 - e^{-\lambda_r t_r}), \tag{35}$$
$$a_n = r_n e^{-\lambda_a t_a} + a_\infty(1 - e^{-\lambda_a t_a}), \tag{36}$$

where r_n and a_n represent $b(t)$ at $t = nt_a$ and $t = n(t_a + t_r)$ respectively, and

$$\lambda_r = K_r D + L_r, \tag{37}$$

$$r_\infty = \frac{K_r D}{K_r D + L_r}, \tag{38}$$

$$\lambda_a = K_a D + L_a, \tag{39}$$

$$a_\infty = \frac{K_a D}{K_a D + L_a}, \tag{40}$$

being K_r, L_r, K_a and L_a the apparent binding rates associated with the two intervals (activation and recovery). r_n and a_n follow an exponential trajectory. Focusing on r_n, its behaviour can be described as (by combining equation (35) and (36)):

$$r_n = r_{ss} + (r_0 - r_{ss})e^{-n\lambda}, \tag{41}$$

where:

$$r_{ss} = a_\infty + \gamma_r(r_\infty - a_\infty), \tag{42}$$

$$\lambda = \lambda_a t_a + \lambda_r t_r, \tag{43}$$

$$\gamma_r = \frac{1 - e^{-\lambda_r t_r}}{1 - e^{-\lambda}}. \tag{44}$$

It is possible to estimate the fraction of bound channels by observing how the blockage modifies the channels conductance and the sodium channel current I. Given a membrane potencial V

$$I = g(1 - b)V. \tag{45}$$

Since it is difficult to measure I, the maximum first derivative of the membrane potencial $\dot{V}_{max}$ can be used, because it is theoretically proportional to I. Several measurements of $\dot{V}_{max}$ are made just at the beginning of the activation interval. So the measurement sequence will follow an exponential pattern

proportional to the one expressed in (41):

$$\dot{V}_{max,n} = \dot{V}_{ss} + (\dot{V}_0 - \dot{V}_{ss})e^{-n\lambda}. \tag{46}$$

The proportionality rate between $\dot{V}_{max}$ and $b_n = r_n$ can be determined by:

$$b_n = 1 - \frac{\dot{V}_{max,n}}{\dot{V}_c}, \tag{47}$$

being $\dot{V}_c$ the observation of $\dot{V}_{max}$ made in the absence of the drug.

4.2   Robust identification

The data from Starmer (1988) will be used in the identification process. A concentration $D = 16\mu M$ of cibenzoline, $t_a = 1ms.$ and $t_r = 0.5, 1.0, 1.5, 2.0$ and $2.5s.$ are used in different experiments which are shown in figure 8. The data is divided into two groups, Ω_{ide} which contains data for the identification process of the FPS (columns 1 and 5) and Ω_{val} which contains data for the FPS validation (columns 2, 3 and 4)

Figure 9 shows the I/O model structure where:

- $\dot{V}_{max}$ is the maximum first derivative of the membrane potential in V/s
- t_a and t_r are the activation and recovery intervals respectively.
- D is the drug concentration in M (mol/l).
- $\theta = [K_a, L_a, K_r, L_r]^T$ are the model parameters. The model is non-linear with respect to θ.

It is necessary to know $\dot{V}_o$ to determine $\dot{V}_{max}$. So $\dot{V}_o = \dot{V}_{max}(0)$ is assumed. Both ∞-norm $N_1(\theta)$ and absolute norm $N_2(\theta)$ are simultaneously used with
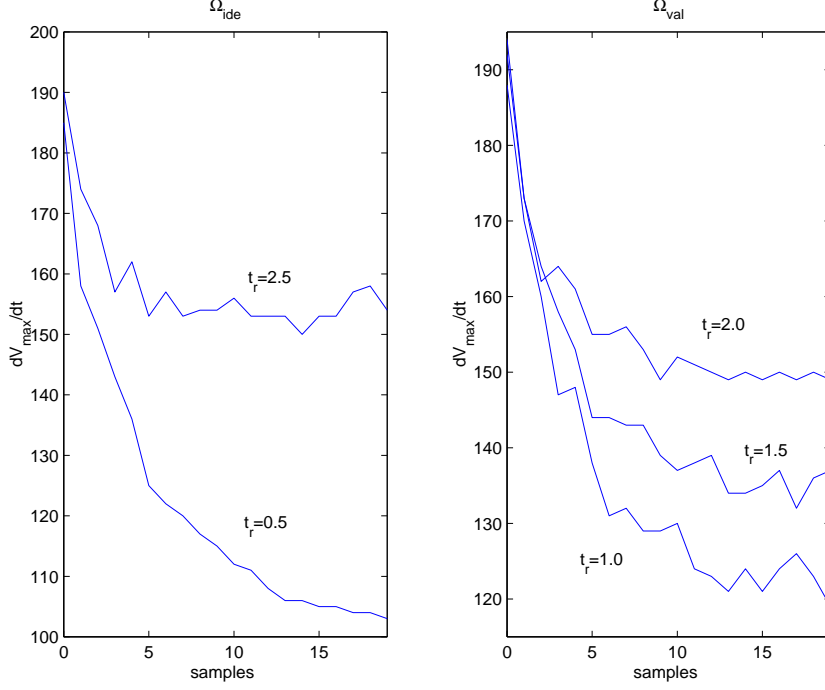
Figure 8. Representation of identification data. Two groups are shown, Ω_{ide} ($t_r = 0.5$ and $t_r = 2.5$) and Ω_{val} ($t_r = 1.0$, $t_r = 1.5$ and $t_r = 2.0$).
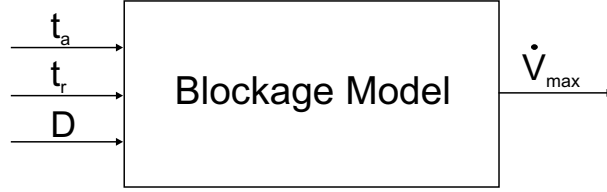
Figure 9. I/O model structure.

the Ω_{ide} data to determine the FPS_{ide}.

$$N_1(\theta) = \|\mathbf{e}(\theta, \Omega_{ide})\|_\infty \tag{48}$$

$$N_2(\theta) = \frac{\|\mathbf{e}(\theta, \Omega_{ide})\|_1}{N_{\Omega_{ide}}} \tag{49}$$

To select the norm bounds η_1 and η_2 and to ensure that it is possible to validate FPS_{ide} with Ω_{val}, the Pareto Front information from following multiobjective optimization problem is considered:

$$\min_{\theta \in D} \mathbf{J}(\theta) = \{N_3, N_2, N_4\}, \tag{50}$$

28

where [13] :

$$N_3(\theta) = \|\mathbf{e}(\theta, \{\Omega_{ide}, \Omega_{val}\})\|_\infty \tag{51}$$

$$N_4(\theta) = \frac{\|\mathbf{e}(\theta, \Omega_{val})\|_1}{N_{\Omega_{val}}} \tag{52}$$

Figure 10 shows the Pareto Front corresponding to the projection optimum models $\hat{\mathbf{\Theta}}_P^*$.
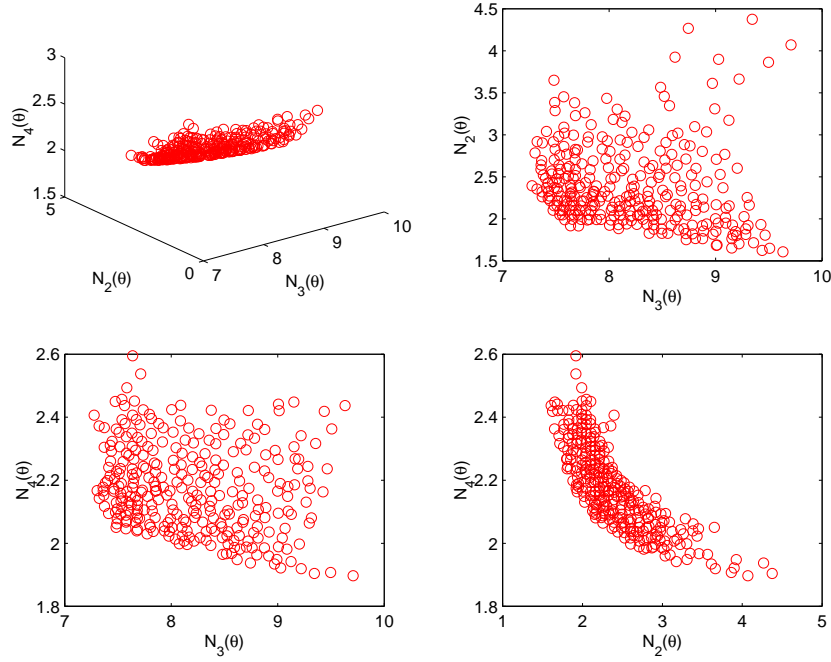
Figure 10. On the top right the $\mathbf{J}(\hat{\mathbf{\Theta}}_P^*)$ Pareto Front and its projection on different planes (the rest of figures).

Bounds $\eta_1 = 8$ and $\eta_2 = 2.5$ are selected from the Pareto Front analysis to hold the FPS_{ide} model prediction errors not greater than $8V/s$ and their average values not greater than $2.5V/s$. So $\hat{\Theta}_{Pr} \neq \emptyset$ and $FPS \neq \emptyset$. The FPS_{ide}

---

[13] Since $\|\mathbf{e}(\theta, \{\Omega_{ide}, \Omega_{val}\})\|_\infty^\mathbf{w} = \max(\|\mathbf{e}(\theta, \Omega_{ide})\|_\infty^\mathbf{w}, \|\mathbf{e}(\theta, \Omega_{val})\|_\infty^\mathbf{w})$ it is sufficient to use the $N_3(\theta)$ norm with data Ω_{ide} and Ω_{val} simultaneously to determine a maximum bound on the ∞-norm with data Ω_{ide} and Ω_{val} separately.

validation is also ensured since there are FPS_{ide} models consistent with Ω_{val} and bounds η_1 and η_2 (see the projection $(N_3(\theta), N_4(\theta))$ of $\mathbf{J}(\hat{\boldsymbol{\Theta}}_P^*)$).

The FPS_{ide} is determined next by $\epsilon - GA$ with the following parameters:

- Searching space: $K_a \in [1e^{-4} \ldots 1e^8]\frac{M}{s}$, $L_a \in [1e^{-4} \ldots 1e^3]\frac{1}{s}$, $K_r \in [1e^{-4} \ldots 1e^4]\frac{M}{s}$ and $L_r \in [1e^{-4} \ldots 0.5]\frac{1}{s}$.

- $t_{max} = 40000$ and $\epsilon = [1e6, 10, 100, 0.005]$ so the grid contains 100 divisions per dimension.

- $Nind_P = 100$, $Nind_G = 4$, $P_{c/m} = 0.1$, $d_{ini} = 0.25$, $d_{fin} = \beta_{fin} = 0.1$ and $\beta_{ini} = 10$.

- The parameter $\delta(t)$ is tuned as

$$\delta(t) = \delta'(t) \cdot \bar{J}, \tag{53}$$

in order to be useful for other optimization problems, where $\bar{J}$ is the J average for all the individuals inserted in the population $P(t)$ during the optimization process. An average estimation of function J is obtained and δ is related to the optimization problem [14]. $\delta'(t)$ is determined by:

$$\delta'(t) = \frac{\delta_{ini}}{\sqrt{1 + \left(\left(\frac{\delta_{ini}}{\delta_{fin}}\right)^2 - 1\right)\frac{t}{(t_{max}-1)}}}, \tag{54}$$

with $\delta_{ini} = 0.1$ and $\delta_{fin} = 0.01$.

Figure 11 shows the ϵ-GA optimization process result, i.e. FPS_{ide}^*. The FPS_{ide} has been characterized by 927 models and the $J(\partial FPS_{ide}^*)$ average is 0.00942, which shows good algorithmic convergence (the ideal $J(\partial FPS_{ide}^*)$ average would be 0). The FPS_{ide}^* is validated because it contains 19 models consistent

---

[14] Only those values inserted in $P(t)$ lower than $\bar{J}$ are taken into account to ensure that $\delta(t)$ never increases.
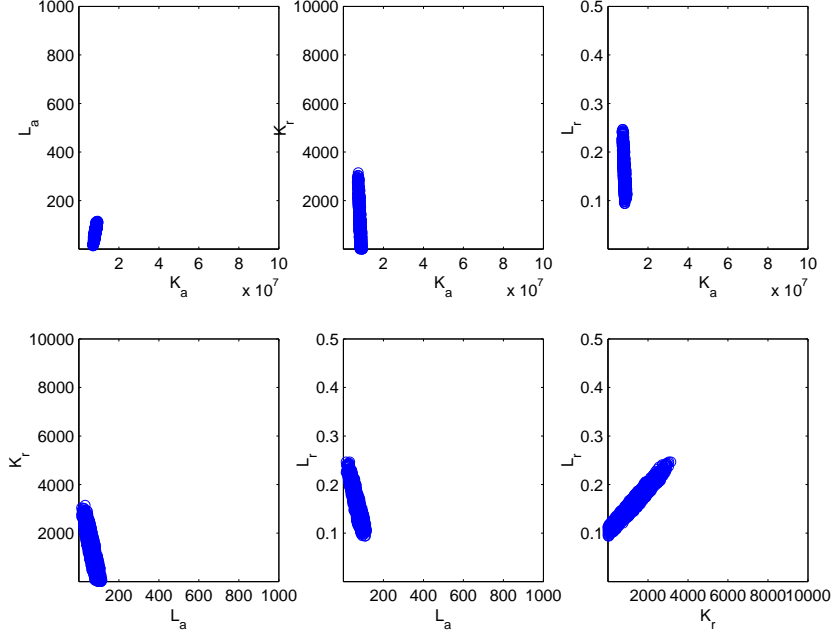
with Ω_{val}.

Figure 11. The FPS^*_{ide} model projections inside the searching space.

Figure 12 shows the Ω_{ide} data, and the envelope generated by the FPS^*_{ide}, whereas the Ω_{val} data and the envelope from FPS^*_{ide} is shown in figure 13.

Once the FPS^*_{ide} has been validated, the restricted interpolatory projection nominal model $\hat{\theta}^*_{pi}$ is determined by first obtaining the worst case model $\hat{\theta}^*_c$.

$$\hat{\theta}^*_c = [8.252e6, 67.28, 1406.80, 0.1657]^T, \tag{55}$$
$$\hat{\theta}^*_{pi} = [8.493e6, 82.56, 11.38, 0.1142]^T. \tag{56}$$

With these models, the following N_1, N_2, N_3 and N_4 norm values are obtained:

$$N_1(\hat{\theta}^*_c) = 6.74V/s, \ \ N_1(\hat{\theta}^*_{pi}) = 7.41V/s \tag{57}$$
$$N_2(\hat{\theta}^*_c) = 2.21V/s, \ \ N_2(\hat{\theta}^*_{pi}) = 2.37V/s \tag{58}$$
$$N_3(\hat{\theta}^*_c) = 8.83V/s, \ \ N_3(\hat{\theta}^*_{pi}) = 7.41V/s \tag{59}$$
$$N_4(\hat{\theta}^*_c) = 3.12V/s, \ \ N_4(\hat{\theta}^*_{pi}) = 2.24V/s \tag{60}$$

31
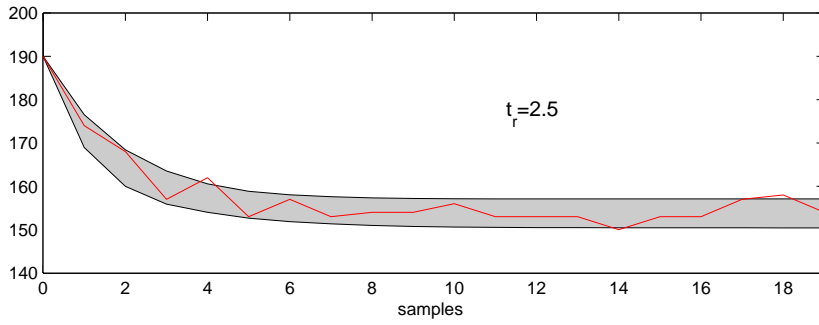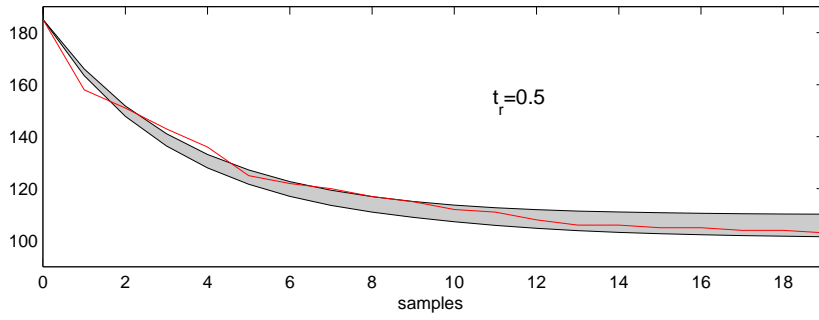
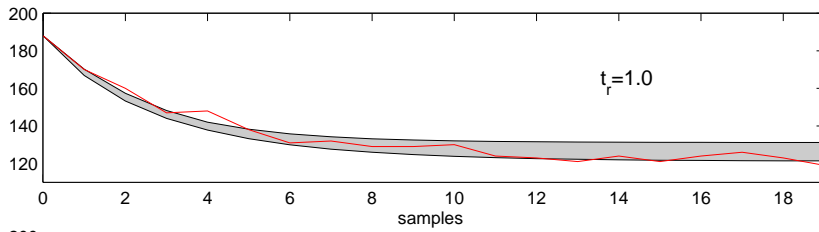Figure 12. $y_{ide}(t)$ and the FPS^*_{ide} model envelopes.
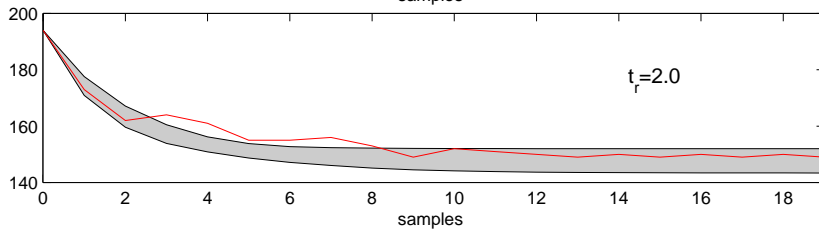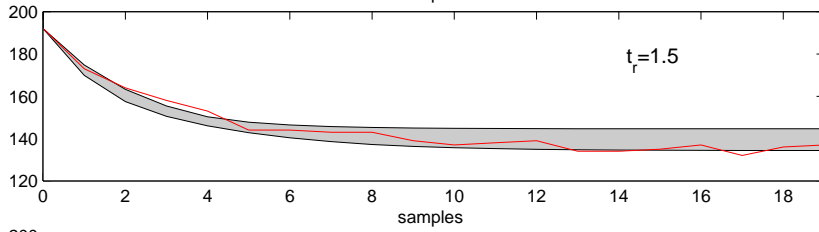
Figure 13. $y_{val}(t)$ and the FPS^*_{ide} model envelopes.

32

so $\hat{\theta}_c^*$ is not validated (since $N_3(\hat{\theta}_c^*) > 8V/s$ and $N_4(\hat{\theta}_c^*) > 2.5V/s$), whereas $\hat{\theta}_{pi}^*$ is (since to $FPS = FPS_{ide} \bigcap FPS_{val}$) and, therefore, its selection as nominal model is more appropriate.

Figure 14 shows the location of model $\hat{\theta}_c^*$, $\hat{\theta}_{pi}^*$ together with the FPS_{ide}^* and the restricted projection models $\hat{\Theta}_{Pr}^*$.
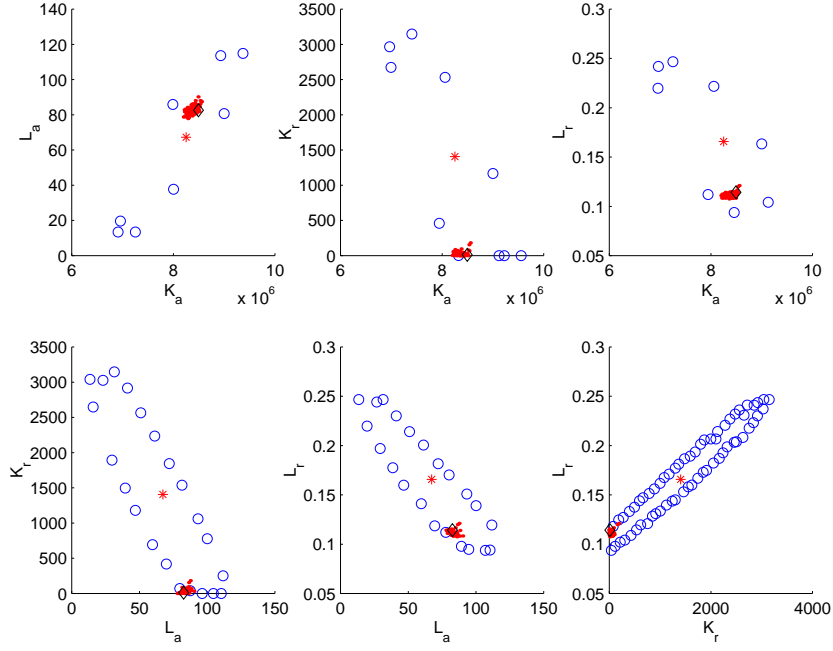
Figure 14. Optimum nominal models. (*) $\hat{\theta}_c^*$, (◇) $\hat{\theta}_{pi}^*$ and (·) $\hat{\Theta}_{Pr}^*$.

The following nominal model is identified by Starmer (Starmer, 1988) using non-linear programming

$$\hat{\theta}_{Starmer} = [7.49e6, 43.19, 1.439, 0.1148]^T \tag{61}$$

with data $\{\Omega_{ide}, \Omega_{val}\}$, to minimize

$$N_5(\theta) = \frac{\|\mathbf{e}(\theta, \{\Omega_{ide}, \Omega_{val}\})\|_2}{N_{\{\Omega_{ide}, \Omega_{val}\}}} \tag{62}$$

33

and comparing the following result with $\hat{\theta}_c^*$ and $\hat{\theta}_{pi}^*$

$$N_5(\hat{\theta}_c^*) = 0.354V/s, \; N_5(\hat{\theta}_{pi}^*) = 0.292V/s, \; N_5(\hat{\theta}_{Starmer}) = 0.41V/s. \quad (63)$$

As can be seen, the $\hat{\theta}_{Starmer}$ result is worse than $\hat{\theta}_c^*$ and $\hat{\theta}_{pi}^*$, in spite of the fact that the final results have not been estimated by minimizing $N_5(\theta)$ [15]. It is proven that $\hat{\theta}_{Starmer}$ is not a global minimum. In Martínez (2006) a minimum of $N_5(\theta)$ is obtained [16]

$$\hat{\theta}_{N_5} = [7.02e6, 663.577, 215.06, 0.1046] \quad (64)$$

$$N_5(\hat{\theta}_{N_5}) = 0.263V/s, \; N_3(\hat{\theta}_{N_5}) = 10.125V/s \quad (65)$$

whose result is the best with respect to N_5, although it does not belong to the FPS since the result of N_3 norm exceeds the desired bounds of $8V/s$.

5 Conclusions

A methodology, based on a specific genetic algorithm $\epsilon-$GA, has been developed to find the Feasible Parameter Set (FPS) of a non-linear model under parametric uncertainty. That robust identification problem is stated by assuming, simultaneously, the existence of several bounds in identification error. The algorithm presents the following features:

- Assuming parametric uncertainty, many processes can be identified if their outputs can be calculated by model simulation. Differentiability with respect

---

[15] In the same way, the $\hat{\theta}_c^*$ and $\hat{\theta}_{pi}^*$ results are better than $\hat{\theta}_{Starmer}$ for the rest of norms.

[16] A classic GA has been used for the optimization.

to the unknown parameters is unnecessary.

- Because more than one norm is taken into account at the same time, the computational cost is reduced since various FPS_i intersections are implicitly performed.

- Non-convex and even disjoint FPS can be calculated.

- Since FPS is not approximated by either orthotopes, or ellipsoids, a non-conservatism is provided.

An approach has been presented which makes the determination of the η_i bounds for the $N_i(\theta)$ norms easier. It is based on the Pareto Front analysis which is obtained by minimizing the norms simultaneously. So, it is possible to choose η_i to achieve $FPS \neq \emptyset$.

The technique obtains a good approximation of the worst case nominal model θ_c, calculating the Chebyshev centre θ_c^* of FPS^*.

This model, in the same way as the FPS determination, is sensitive to η_i bounds, so an alternative to the Chebyshev centre of FPS^* has been presented, which consists of determining the nearest Pareto point to the Chebyshev centre: $\hat{\theta}_{pi}^*$. This model is optimal from two points of view: the identification error (since it belongs to the Pareto Front and so is a projection model); and estimation in the searching space (since it is the nearest projection model to the Chebyshev centre of FPS).

The RI application with real data about cardiac cell blockage caused by the cibenzoline shows the flexibility and power of the proposed RI methodology. It is evidence that the proposed nominal model $\hat{\theta}_{pi}^*$ belongs to the FPS and it is a good trade-off solution for the simultaneously stated norms.

Acknowledgments

References

Bai, E., Ye, Y., Tempo, R., 1999. Bounded error parameter estimation: A survey. IEEE Transaction on Automatic Control. 44(6), 1107–117.

Belforte, G., Bona, B., Cerone, V., 1990. Parameter estimation algorithms for a set-membership description of uncertainty. Automatica. 26(5), 887–898.

Cardona, K.E., 2005. Modelización del comportamiento eléctrico del fármaco antiarrítmico clase ib- lidocaína en células del miocardio en cobaya y conejo. Technical report. Departament of Electronic Engineering, Polytechnic University of Valencia.

Chisci, L., Garulli, A., Vicino, A., Zappa, G., 1998. Block recursive parallelotopic bounding in set membership identification. Automatica. 34(1), 15–22.

Fogel, E., Huang, F., 1882. On the value of information in system identification-bounded noise case. Automatica. 18(12), 229–238.

Garulli, A., Kacewicz, B., Vicino, A., Zappa, G., 2000. Error bounds for conditional algorithms in restricted complexity set membership identification. IEEE Transaction on Automatic Control. 45(1), 160–164.

Herrero, J.M., Blasco, X., Martínez, M., Ramos, C., 2005. Nonlinear robust identification using multiobjective evolutionary algorithms. Lecture Notes in Computer Science. 3562, 231–241. Springer-Verlag.

Herrero J.M., 2006. Robust identification of non-linear systems using evolutionary algorithms. Ph.D. thesis, Polytechnic University of Valencia, Spain.

Keesman, K.J., Stappers, R., 2004. Nonlinear set-membership estimation: A support vector machine approach. J. Inv. Ill-Posed Problems. 12(1), 27–41.

Ljung, L., 1999. System Identification, Theory for the user, 2nd ed. Prentice-Hall.

Martínez, M., 2006. Estimación de parámetros en modelos de bloqueo de fármacos. Technical report. Department of Systems Engineering and Control, Polytechnic University of Valencia.

Milanese, M., Vicino, A., 1991. Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. Automatica. 27(6), 997–1009.

Norton, J.P., 1987. Identification and application of bounded-parameter models. Automatica. 23(4), 497–507.

Starmer, C.F., 1988. Characterizing activity-dependent processes with a piecewise exponencial model. Biometrics. 44, 549–559.

Walter, E., Piet-Lahanier, H., 1990. Estimation of parameter bounds from bounded-error data: A survey. Mathematics and computers in simulation. 32, 449–468.

Walter, E., Pronzalo, L., 1991. Recursive robust minmax estimation for models linear in their parameters. In Proc. of the IFAC Symposium on Identification and system parameter estimation. 1, 763–768.

Walter, E., Pronzalo, L., 1997. Identification of parametric models from experimental data, Springer.

Walter, E., Kieffer, M., 2003. Interval analysis for guaranteed nonlinear parameter estimation. In Proc. of the 13th IFAC Symposium on System Identification.

List of Figures