Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC-PapersOnLine 49-26 (2016) 077–082

# Parameter identification in synthetic biological circuits using multi-objective optimization ⋆

**Y. Boada** * **A. Vignoni** ** **G. Reynoso-Meza** *** **J. Picó** *

* *I.U. de Automática e Informática Industrial (ai2), Universitat Politècnica de Valencia, 46022, Camino de Vera S/N, Valencia, Spain. (e-mail: {yaboa,jpico}@upv.es)*

** *Center for Systems Biology Dresden and Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauer str. 108, 01307 Dresden, Germany. (e-mail: vignoni@mpi-cbg.de)*

*** *Industrial and Systems Engineering Graduate Program (PPGEPS), Pontifícia Universidade Católica do Paraná, Imaculada Conceição, 1155, 80215-901 Curitiba, PR, Brazil. (e-mail: g.reynosomeza@pucpr.br)*

**Abstract:** Synthetic biology exploits the of mathematical modeling of synthetic circuits both to predict the behavior of the designed synthetic devices, and to help on the selection of their biological components. The increasing complexity of the circuits being designed requires performing approximations and model reductions to get handy models. Parameter estimation in these models remains a challenging problem that has usually been addressed by optimizing the weighted combination of different prediction errors to obtain a single solution. The single-objective approach is inadequate to incorporate different kinds of experiments, and to identify parameters for an ensemble of biological circuit models.
We present a methodology based on multi-objective optimization to perform parameter estimation that can fully harness to ensembles of local models for biological circuits. The methodology uses a global multi-objective evolutionary algorithm and a multi-criteria decision making strategy to select the most suitable solutions. Our approach finds an approximation to the Pareto optimal set of model parameters that correspond to each experimental scenario. Then, the Pareto set was clustered according to the experimental scenarios. This, in turn, allows to analyze the sensitivity of model parameters for different scenarios. Finally, we show the methodology applicability through the case study of a genetic incoherent feed-forward circuit, under different concentrations of the inducer input signal.



*Keywords:* Biological circuits, Kinetic parameters, Parameter identification, Multi-objective optimization

## 1. INTRODUCTION

Biological circuits in synthetic biology are usually modeled with systems of ordinary differential equations (ODEs) in order to describe the time-evolution of the involved species concentrations, like mRNA or proteins. Starting from set of biochemical reactions for the circuit, dynamic balances for the biochemical species can be obtained using well established methods, like the mass-action kinetics formalism (Chellaboina et al., 2009; Picó et al., 2015). The resulting dynamic models are high dimensional, even for small circuits. Therefore, model reduction is carried out exploiting the different time scales present in the system (Prescott and Papachristodoulou, 2014). The resulting model depends on several parameters. On the one hand, among these parameters, some of them like binding and unbinding rates, or production and degradation coefficients have a physical meaning. On the other hand, it is possible to obtain other parameters from model reductions and/or approximations, but making more difficult the biological interpretation of its values. Nevertheless, in general both kind of parameters have unknown values for a particular model. Thus, the problem of parameter identification, that is the indirect determination of the unknown parameters from measurements of other quantities, is a key issue in computational and systems and synthetic biology (Lillacci and Khammash, 2010). Accurate parameter identification is crucial whenever one wants to obtain quantitative, or even qualitative information from the models (Lillacci and Khammash, 2010). Recently, much attention has been given to this problem in the systems biology community, using optimization techniques such as linear and nonlinear least squares (Mendes and Kell, 1998), genetic algorithms (Srinivas and Patnaik, 1994), and evolutionary computation (Ashyraliyev et al., 2008; Moles et al., 2003). Evolutionary computation is one of the suggested optimization techniques for the large parameter estimation problems present in systems and synthetic biology (Moles et al., 2003).

Parameter estimation in nonlinear dynamic models remains a very challenging inverse problem due to its nonconvexity, and

⋆ This work is partially supported by Spanish government and European Union (FEDER-CICYT DPI2011-28112-C04-01, and DPI2014-55276-C5-1). Y.B. thanks grant FPI/2013-3242 of Universitat Politècnica de València and Becas Iberoamérica of Santander Group, Spain 2015. G.R.M. thanks the partial support provided by the postdoctoral fellowship BJT-304804/2014-2 from the National Council of Scientific and Technologic Development of Brazil. A.V. thanks the Max Planck Society, the CSBD and the MPI-CBG. We are grateful to Dr. C.Baüerl and Dr. D. Provencio at the $SB_2$CLab for their help in plasmid construction and getting experimental data. Also to Dr. V. Monedero at IATA-CSIC for allowing us to use the POLARstar plate reader at his lab.

2405-8963 
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2016.12.106

ill-conditioning caused by over-parametrization, experimental measurement errors, data scarcity and uncertainty (Gábor and Banga, 2015; Kaltenbach et al., 2009). Moreover, for nonlinear models, the amount of information collected from an experiment may strongly depend on the true value of the parameters (Pronzato and Pázman, 2013). One of the main problems associated with standard optimization methods is that they may not perform well in the case of significant difference in the system response to different inputs. The main reason of this deteriorated performance is that all these identification methods rely on single objective optimization and try to find only one solution (i.e. only one value for each parameter), the *best fit*. This best solution can be good for one set of experiments and bad for others, or it can be acceptable for all the experiments but not really good for any one.

Several approaches have been proposed to tackle these problems. Among them, ensembles of local models have received much attention in the last years, when a single set of parameters is not appropriate for all experimental scenarios. In Steuer et al. (2006), local linear models at each point in parameter space where used to circumvent lack of knowledge about the structure of kinetics by a parametric representation of the Jacobian matrix. Then, the authors used the ensemble of models to elucidate the parameter regions associated with experimentally observed specific dynamical behaviors. A similar approach was used by Samee et al. (2015). Ensembles of models, i.e. sets of models with different structures and/or parameter values have also been used in Villaverde et al. (2015), where the final prediction is obtained from a consensus one among the models.

In this work, we propose a methodology based on using multi-objective optimization design (MOOD) to perform parameter identification leading to nonlinear local models of biological circuits. The methodology uses a *global* multi-objective evolutionary algorithm (MOEA) and a multi-criteria decision making (MCDM) strategy to select the most suitable solutions (Reynoso-Meza et al., 2014). Although the identification problem itself can be naturally expressed as a multi-objective problem (MOP), this approach has been seldom used (Velasco-Carrau et al., 2015; Bonilla-Petriciolet et al., 2013). Our approach uses a MOEA to find the best approximation to the Pareto set (see Reynoso-Meza et al. (2010) for characterization details and benchmarks of the algorithm) of model parameters that correspond to each experimental scenario. The Pareto set together with the Pareto front regions are correlated with the experimental scenarios using kmeans clustering. This, in turn, allows to perform a MCDM and analyze which model parameters vary to explain each scenario. To show the applicability of the methodology we performed the multi-objective optimization based identification on a well-known biological circuit, a genetic incoherent feed-forward loop showing adaptive behavior, under different concentrations of the inducer input signal.

The rest of the paper is organized as follows: Section 2.1 describes the biological circuit used, the experimental implementation, and its model. In section 2.2 the proposed methodology is described. The results achieved are shown in Section 3 where the main findings are presented and, finally conclusions are drawn in the last section.

## 2. MATERIALS AND METHODS

### 2.1 Incoherent type 1 feed-forward loop (I1-FFL)

Adaptation is an important property of biological systems, linked to homeostasis (Alon, 2006). The incoherent type 1 feed-forward loop (I1-FFL), depicted in Fig. 1, is one the most common network motifs showing adaptation. Different implementations are possible, including enzyme reaction networks (Ma et al., 2009; Chiang et al., 2014), gene networks (Basu et al., 2004) and *in vitro* transcriptional networks (Kim et al., 2014).

*Experimental implementation* Following the implementation in (Basu et al., 2004), we engineered and implemented the circuit in the lab using components taken from the Lux operon in the *V. fisheri* quorum sensing system, the lambda cI repressor and a green fluorescent protein as a reporter.

Figure 1 depicts the gene synthetic circuit. The extracellular AHLe acts as input to the circuit. The protein LuxR binds to the intracellular AHL, forming a monomer $\mathrm{LuxR} \cdot \mathrm{AHL}$. This one dimerizes forming $(\mathrm{LuxR} \cdot \mathrm{AHL})_2$. The dimer $(\mathrm{LuxR} \cdot \mathrm{AHL})_2$ is the transcription factor that directly activates expression of the gene *gfp*, and indirectly represses it via activation of the repressor cI. As a result, when the signal AHL causes the LuxR node to assume its active conformation, GFP is produced. After some time cI accumulates and forms the dimer $(\mathrm{cI})_2$, which eventually attains the repression threshold for the hybrid promoter of *gfp* gene and makes the level of GFP decrease.

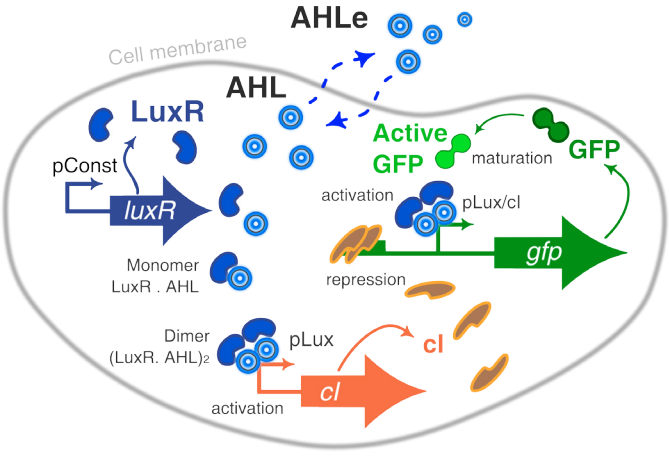


Fig. 1. Representation of a cell incorporating the engineered incoherent feed-forward loop synthetic circuit.

Two plasmids are used to build the circuit. On the one hand, in the plasmid pCB14mut, the gene coding for the protein LuxR (BBa_C0062) is constitutively expressed under the control of a medium strength promoter (BBa_J23106) and a strong RBS (BBa_B0034). Also, in the same plasmid, a pLux/cI hybrid promoter (BBa_K415032) drives the expression of GFP (BBa_K082003) with a strong RBS (BBa_B0034). This two cassettes are placed in a pBR322 plasmid backbone. On the other hand, the plasmid pCB11a contains the gene *cI* (BBa_K327018) controlled by the pLux repressible promoter (BBa_R0062) and a mild ribosome binding site (RBS part BBa_B0033) in the pACYC184 plasmid backbone. All parts were taken from the Registry of Standard Biological Parts (Biobrick Foundation, 2006) and cloned using the 3 Antibiotic As-

sembly method from Biobrick fundation (Biobrick Foundation, 2006). All coding sequences are followed by the terminator BBa_B0015.

The following experiments were performed. *E.coli* ells (Top 10, NEB) carrying the pCB11a and pCB16mut plasmids were grown over night in LB medium with the appropriate antibiotics. Then, 96 well-plates were inoculated at $\mathrm{OD}_{600} \approx 0.025$ and incubated to reach an optical density of $\mathrm{OD}_{600} \approx 0.2$. At this point, selected wells were induced with appropriate concentrations of AHL (N-3-Oxohexanoyl-L-homoserine lactone, Santa Cruz Biotecnology Catalog Number SC205396) and incubated for 200 minutes. Measurements were taken with a POLARstar Omega plate reader (BMG Labtech GmBh) with the following protocol: 2' shaking, OD measurement, then 15" pause and Fluorescent measurement. Each condition was performed in 4 replicates, and in 2 different days making a total of 8 datasets for each condition. From each experiment, we have 6 conditions (ranging from 0 nM to 55 nM AHL), 8 datasets, absorbance and fluorescence measurements every 5 minutes during 200 minutes of incubation after induction.

*Mathematical model* We use an ODE model adapted from (Boada et al., 2016) with twelve states. The species in the model are mLuxR, LuxR, AHL, $(\mathrm{LuxR}\cdot\mathrm{AHL})_2$, mcI, cI, $(\mathrm{cI})_2$, mGFP, GFP, aGFP, AHLe and the number of cells in the culture respectively (see Table 1). The system of ODEs describing the dynamics of the circuit is:

$$
\begin{aligned}
\dot{x}_1 &= k_{m_{\mathrm{luxR}}} C_{p1} - d_{m_{\mathrm{luxR}}} x_1 - \mu x_1 \\
\dot{x}_2 &= k_{p_{\mathrm{luxR}}} x_1 - k_2 x_2 x_3 + k_{-2} M - d_{\mathrm{LuxR}} x_2 - \mu x_2 \\
\dot{x}_3 &= -k_2 x_2 x_3 + k_{-2} M + k_d x_9 - k_{-d} x_3 - d_{\mathrm{AHL}} x_3 - \mu x_3 \\
\dot{x}_4 &= k_3 M^2 - k_{-3} x_4 - \mu x_4 \\
\dot{x}_5 &= k_{m_{\mathrm{cI}}} C_{p2} \frac{x_4}{\gamma_1 + x_4} - d_{m_{\mathrm{cI}}} x_5 - \mu x_5 \\
\dot{x}_6 &= k_{p_{cI}} x_5 - 2k_4 {x_6}^2 + 2k_{-4} x_7 - d_{cI} x_6 - \mu x_6 \\
\dot{x}_7 &= k_4 {x_6}^2 - k_{-4} x_7 - \mu x_7 \\
\dot{x}_8 &= \frac{k_{m_{\mathrm{GFP}}} C_{p1} (x_4 + c_0\gamma_2 + \beta_2\gamma_5 x_4 x_7)}{\gamma_2 + \gamma_3 x_4 + \gamma_4 x_7 + \gamma_5 x_4 x_7} - d_{m_{\mathrm{gfp}}} x_8 - \mu x_8 \\
\dot{x}_9 &= k_{p_{\mathrm{gfp}}} x_8 - \frac{\ln 2}{k_{\mathrm{mat}}} x_9 - d_{\mathrm{GFP}} x_9 - \mu x_9 \\
\dot{x}_{10} &= \frac{\ln 2}{k_{\mathrm{mat}}} x_9 - d_{\mathrm{GFP}} x_{10} - \mu x_{10} \\
\dot{x}_{11} &= K_{\mathrm{cells}} (k_{-d} x_3 - k_d x_{11}) - d_{\mathrm{AHLe}} x_{11} \\
\dot{x}_{12} &= \mu x_{12} \left(1 - \frac{x_{12}}{K_{\max}}\right)
\end{aligned}
\tag{1}
$$

with $M = \frac{\sqrt{(d_M + k_{-2} + \mu)^2 + 8k_3(k_2 x_2 x_3 + 2k_{-3} x_4)}}{4k_3} + \frac{d_M + k_{-2} + \mu}{4k_3}$ being the monomer algebraic relation and $K_{\mathrm{cells}} = \frac{V_{cell} \cdot x_{11}}{V_{med}}$ the volumes relationship in order to correct concentrations outside the cells. From the experimental set up we have $V_{cell} = 10^{-15}\mathrm{L}$ (volume of an *E.coli* cell) and $V_{med} = 180\mu\mathrm{L}$ (the culture medium used in the 96-well plate reader).

Model (1) was obtained after a reduction process from the original equations resulting from the application of mass-action kinetics on the biochemical reactions. During this reduction process some simplifications are used, both by assuming fast dynamics for the heterodimerization process yielding $(\mathrm{LuxR}\cdot\mathrm{AHL})_2$, and by considering a quasi-empirical simplification for the activation function of the hybrid promoter. This simpler model allows easier theoretical and computational analysis. However, given the nonlinear behavior of the system, it is not possible to pretend that a single set of parameters will be appropriate for all experimental scenarios anymore.

Table 1. List of variables used in the model

| Variable | Description | Units |
|---|---|---|
| $x_1$ | LuxR mRNA | nM |
| $x_2$ | LuxR protein | nM |
| $x_3$ | Acyl-homoserine lactone (AHL) | nM |
| $M$ | LuxR·AHL monomer | nM |
| $x_4$ | $(\mathrm{LuxR}\cdot\mathrm{AHL})_2$ dimer | nM |
| $x_5$ | cI mRNA | nM |
| $x_6$ | cI protein | nM |
| $x_7$ | $(\mathrm{cI})_2$ dimer | nM |
| $x_8$ | GFP mRNA | nM |
| $x_9$ | GFP protein | nM |
| $x_{10}$ | Active GFP protein | nM |
| $x_{11}$ | Extracellular AHL | nM |
| $x_{12}$ | Number of cells in the culture | Cells |

Table 2. Parameters for the model

| **Fixed Parameter** | **Description** | **Value** |
|---|---|---|
| $k_{m_{\mathrm{luxR}}}$ | luxR transcription rate | 1 $\mathrm{min}^{-1}$ |
| $k_{p_{\mathrm{luxR}}}$ | LuxR translation rate | 50 $\mathrm{min}^{-1}$ |
| $k_d, k_{-d}$ | AHL diffusion rate | 2 $\mathrm{min}^{-1}$ |
| $k_4, k_{-4}$ | $(\mathrm{cI})_2$ association, dissociation rate | 0.0009, 0.6 $\mathrm{min}^{-1}$ |
| $C_{p1}$ | Plasmid pBR322 copy number | 17 |
| $C_{p2}$ | Plasmid pACYC184 copy number | 15 |
| $\gamma_2$ | Hybrid pLuxR/cI promoter coefficient | 0.02 nM |
| $d_{m_{\mathrm{luxR}}}, d_{m_{\mathrm{cI}}}, d_{m_{\mathrm{gfp}}}$ | mRNAs degradation rates | 0.23 $\mathrm{min}^{-1}$ |
| $d_{\mathrm{LuxR}}$ | LuxR degradation rate | 0.0174 $\mathrm{min}^{-1}$ |
| $d_{\mathrm{AHL}}, d_{\mathrm{AHLe}}$ | AHL degradation rates | 0.01 $\mathrm{min}^{-1}$ |
| $d_M$ | Monomer degradation rate | 0.0174 $\mathrm{min}^{-1}$ |
| $K_{\max}$ | maximum growth capacity | $1.62 \cdot 10^8$ cells |
| $\mu$ | Specific growth rate | 0.028 $\mathrm{min}^{-1}$ |
| **Unknown Parameter** | **Description** | **Range of values** |
| $d_{\mathrm{cI}}, d_{\mathrm{GFP}}$ | cI, GFP degradation rate | [0.01 0.3] $\mathrm{min}^{-1}$ |
| $\gamma_1$ | pLux Promoter Hill constant | [50 100] nM |
| $\gamma_3$ | Hybrid pLuxR/cI promoter coefficient | [0.0001 0.5] |
| $\gamma_4$ | Hybrid pLuxR/cI promoter coefficient | [0.0005 5] |
| $\gamma_5$ | Hybrid pLuxR/cI promoter coefficient | [1 100] |
| $k_{p_{\mathrm{cI}}}, k_{p_{\mathrm{gfp}}}$ | cI, GFP translation rate | [1 60], [1 100] $\mathrm{min}^{-1}$ |
| $c_0$ | Hybrid promoter basal expression | [0 0.01] |
| $\beta_2$ | Hybrid promoter leakiness | [0 0.01] |
| $k_{m_{\mathrm{cI}}}, k_{m_{\mathrm{gfp}}}$ | cI, gfp transcription rate | [0.1 75], [0.1 25] $\mathrm{min}^{-1}$ |
| $k_{-2}, k_{-3}$ | Monomer and dimer dissociation rate | [0.05 0.3], [0.1 1] $\mathrm{min}^{-1}$ |
| $k_2, k_3$ | Monomer and dimer association rate | [0.0006 0.06] $\mathrm{min}^{-1}$ |
| $k_{\mathrm{mat}}$ | GFP maturation time | [20 120] min |

The model has 35 parameters. Out of them, 18 are know from the literature and were kept fixed (see Table 2).

### 2.2 Multi-objective optimization approach

In order to successfully implement this approach, at least three fundamental steps are required (Miettinen et al., 2008): the multi-objective problem (MOP) definition, the optimization process, and the multi-criteria decision making (MCDM) stage. This overall multi-objective optimization design (MOOD) procedure enables to analyze current trade-offs between the objectives to accordingly select a preferable solution (Reynoso-Meza et al., 2014).

*Multi-objective problem definition* At this point the error measures between the experimental data and the model predictions, for each inducer concentration, are formulated as objectives to be optimized. Thus, we optimize the mean square error (MSE) of the active GFP fluorescence for each input

$\text{AHL}_e = \{5, 15, 25, 35, 55\}$ nM. Several samples for each inducer concentration were taken at every $T = 5.65\text{min}$ during approximately 200 min. The second part of Table 2, constitute the parameters of model (1) to be identified and which will be the decision variables of our optimization, denoted $\theta$ in the following.

These design objectives can be expressed with the following indexes:

$$J_{[i=1,\dots,5]}(\theta) = \frac{1}{n}\sum_{q=1}^{n}\frac{1}{m}\sum_{k=1}^{m}\left(x^m_{10_{iq}}(k) - x_{10_{iq}}(kT)\right)^2 \quad (2)$$

where $x^m_{10}$ and $x_{10}$ are the experimental and predicted observations of active GFP at the instant $k$ respectively, $i$ is the design objective for each input value, $n$ is the number of observation copies measured at the point $k$ for the same objective, $m$ is the total number of experimental observations. The input stimulus is applied at $t_0 = 0$.

Finally, we look for the set of values of the 17 decision variables $\theta$ that minimize all objectives $J(\theta)$. These five objectives are in conflict if one tries to identify a single ensemble of parameters. So a trade-off must be reached. Our problem can be formulated as a multi-objective problem:

$$\begin{aligned} \min_{\theta\in\Re^{17}} J(\theta) = \quad & [J_1(\theta), ..., J_5(\theta)] \in \Re^5 \\ \text{subject to:} \quad & \text{equation (1)} \end{aligned} \quad (3)$$

*Multi-objective optimization process* The multi-objective optimization process finds the best parameters $\theta^*_P$ which produce the best Pareto front approximation $J^*_P$. For problems with a large number of decision variables, as our case, it is more efficient to use an appropriate multi-objective optimization algorithm to approximate this solution.

In this work we used a multi-objective evolutionary algorithm based on differential evolution which uses a spherical pruning to approximate the Pareto front. The implementation used in this work is the sp-MODE [1] algorithm, which *i)* improves convergence by using an external file to store solutions and include them in the evolutionary process, *ii)* improves spreading by using the spherical pruning mechanism (Reynoso-Meza et al., 2010), and *iii)* improves pertinency of solutions by means of a basic bound mechanism in the objective space, as described in (Reynoso-Meza et al., 2012).

*Multi-criteria decision making stage* The selection of the preferable solution according to designer's criteria takes place in an *a-posteriori* multi-criteria analysis of the Pareto front approximation. It is desirable that these tools simplify the visualization and the analysis of the trade-off among competing objectives. We use the visualization tool Level Diagrams (Blasco et al., 2008), which has a freely available implementation for designers LD-Tool [2] . LD-Tool allows to correlate design objectives with decision variables by providing two graphs. The first graph contains each objective, where its Y-axis is the p-norm $\|J(\theta)\|_p$ of the objectives vector, and the X-axis corresponds to each objective value $J_i(\theta)$ (see Figure 2). The second graph provided by the LD-Tool shows $\|J(\theta)\|_p$ with respect to each decision variable (graph not shown). Thus, a given solution will have the same $y$-value in all graphs. In addition, the solutions were clustered using the *kmeans* algorithm and all the graphs were colored by the resulting clusters.

[1] Tool available in `http://www.mathworks.com/matlabcentral/fileexchange/39215`

[2] Tool available at `http://www.mathworks.com/matlabcentral/fileexchange/24042`

## 3. RESULTS

We carried out a first optimization of (3) to get an initial estimate of the Pareto front and the unknown parameters in the model. From this preliminary optimization we found appropriate minimum and maximum limits for the MSE of each objective, the so called *pertinency* of each $J_i(\theta)$, which were used to enhance the search of the Pareto front in a narrower region of the parameters space. In both cases, the optimization was done using sp-MODE starting with an initial population of candidate solutions chosen randomly from a uniform distribution in the parameters space.

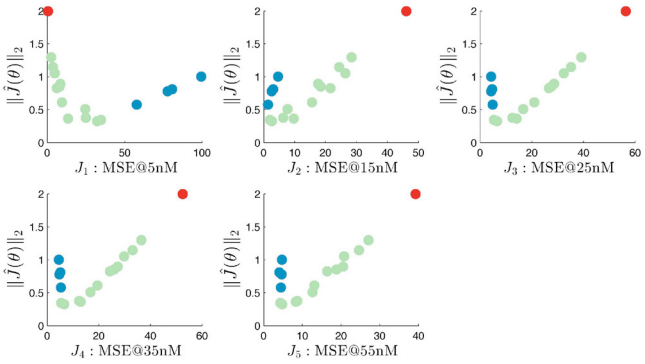


Fig. 2. Identification results: LD-modified representation of the Pareto Front for each objective colored with the three resulting clusters. Each point represent a solution of the MOP.

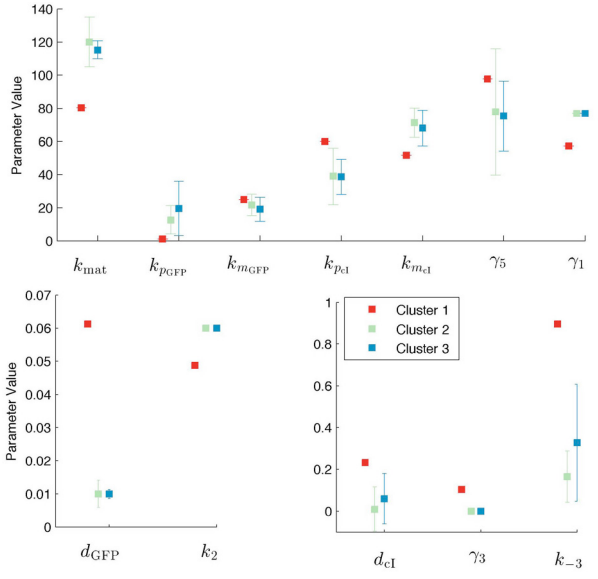


Fig. 3. Identification results: Pareto set representation with the value of the varying parameters in each cluster.

In the next step, an approximation of the Pareto front with 17 solutions is obtained (Figure 2), together with the Pareto set (Figure 3) containing their corresponding parameters. These solutions are classified using the *kmeans* algorithm into three

clusters showing a trade-off between the different objectives corresponding to the different AHL induction levels. This clustering helps to choose best parameters for different cases. For instance, in this example we found that parameters in Cluster 1 (in red) present small errors for the 5 nM induction, but they present larger errors for medium and/or higher induction values. Meanwhile, solutions in clusters 2 and 3 (green and blue respectively) perform better for medium (15 nM) and high (from 25 to 55 nM) inductions than solutions in cluster 1.

The Pareto front analysis shows on the one hand the clasical trade-offs. For instance, the red point, which the best solution for $J_1$ (5 nM) is the worst solution for all the other objectives. Green solutions in cluster 2, are better for $J_2$ (15 nM) but not that good for $J_1$. Finally, blue solutions in cluster 3 are good for $J_3$, $J_4$ and $J_5$ (from 25 to 55 nM) , but not so good for $J_2$, and bad for $J_1$. On the other hand, the Pareto front looks quite similar for objectives $J_3$, $J_4$ and $J_5$, corresponding to inputs ranging from 25 nM to 55 nM respectively, as seen in Figure 2. Moreover the minimum values for these objectives are slightly larger than those obtained for objectives $J_1$ and $J_2$. Recall also that these three objectives all fell within the same cluster. This similarity may be related to the high dependence of promoter activity on the concentration of the AHL induction. Although it has been shown this dependence can saturate and reduce the pLux and pLux/cI promoter activity at levels of AHL $>$ 40 nM, there is no much difference in promoter activity for inductions larger than 20 nM (Egland and Greenberg, 2000). This saturation is observed in the experimental data (see Fig. 4), and captured by the model. Yet, also a delayed peak is observed in the experimental results for large concentrations of the inducer, which is not captured by the model.

Out of the 17 estimated parameters, 5 parameters had the same value in all clusters: $\gamma_4 = 1.42$, $c_0$ = 0.008, $\beta_2 = 0.0014$, $k_3 = 0.0006$ min$^{-1}$ and $k_{-2} = 0.2$ min$^{-1}$. Figure 3 shows the range of values obtained for 12 parameters with different values in each cluster. Notice that parameter values in clusters 2 (15 nM) and 3 (from 25 to 55 nM) are quite alike, and distinctively different from parameter values in cluster 1 (5 nM). As said before, this is a consequence of the saturation with increasing AHL inducer concentration.

A few parameters account for the difference between the model for low inducer concentration (5nM), and for medium-high one (clusters 2 and 3). Note, in particular, the difference in the monomer association rate $k_2$ and the dimer dissociation rate $k_{-3}$. The LuxR-family of transcription factors are believed to be largely disordered (i.e., unfolded) in the monomeric form, becoming folded only upon dimerization in the presence of the inducer (Buchler et al., 2005). Thus, for low values of inducer one may expect larger formation of monomer (larger $k_2$), and dissociation of dimer (lower $k_{-3}$), as seen in Fig. 3.

In addition, the results presented also enables us to see that there is a big difference between values of the degradation rate $d_{GFP}$ in the different clusters. The turnover of GFP clearly decreases in presence of inducer. This may be related to the maximal capacity of the proteases that are present in a bacterial cell (Leveau and Lindow, 2001). Beyond a certain concentration of GFP, their combined proteolytic activity is not enough to reduce the increment of the GFP content.

Finally, from the resulting clusters we selected the median value as a representative of each parameter in each cluster. Fig. 4 shows the comparison between predictions from the model and

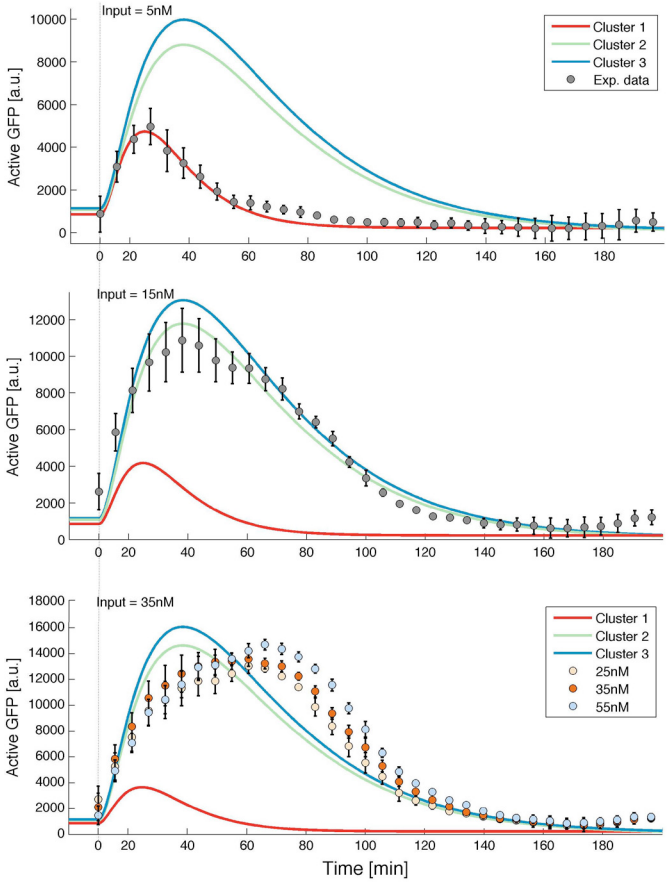


Fig. 4. Comparison of predicted and experimental data for different inductions. Dots correspond to mean values of experimental data (different data sets that the ones used for identification), with its variance as vertical bars. Predictions (continuous line) obtained to the three cluster representatives.

experimental data for different inductions and the identified parameters selected. The validation was performed with data sets not previously used for identification. Note that responses to induction levels from 25nM to 55nM are very similar among them, as it was mentioned in the Pareto front analysis before.

## 4. CONCLUSION

We have proposed a methodology based on multi-objective optimization to perform parameter identification leading to an ensemble of local models for biological circuits. Our approach finds an approximation to the Pareto set of model parameters that correspond to each experimental scenario. Then, the Pareto front regions are associated with different performances for the each experimental scenario. The solutions obtained might be post-processed with a multivariate analysis statistical tool such as *kmeans*, in order to get further insight into the role of the different parameters. This methodology allows to add diverse objectives in the identification process thus enabling to confront different experiments of the same kind.

The proposed methodology benefits was shown for the parameter identification of the incoherent feed-forward loop circuit implementation. In particular, the approach allowed to identify the most relevant parameters for each scenario. We also found that the parameters associated with the formation of the monomer

and dimer, and the degradation rate of GFP vary, as already suggested by other studies. A difference in the peak time between the experimental and estimated circuit output is observed for large concentrations of the inducer. The experimental delayed response may be due to saturation effects not taken into account in the model, and deserve further work. Although the estimated maturation time of GFP was large, it cannot fully account for the peak delay. Some extra dynamics seem to be at play. The overall agreement between experimental and predicted data is remarkable.

## REFERENCES


Alon, U. (2006). *An Introduction To systems biology. Design Principles of Biological Circuits*. Chapman & Hall/ CRC Mathematical and computational Biology Series.

Ashyraliyev, M., Jaeger, J., and Blom, J.G. (2008). Parameter estimation and determinability analysis applied to drosophila gap gene circuits. *BMC Systems Biology*, 2(1), 83.

Basu, S., Mehreja, R., Thiberge, S., Chen, M.T., and Weiss, R. (2004). Spatiotemporal control of gene expression with pulse-generating networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(17), 6355–6360.

Biobrick Foundation (2006). Part registry [online]. `http://partsregistry.org/`. Accesed: 20/02/2015.

Blasco, X., Herrero, J., Sanchis, J., and Martnez, M. (2008). A new graphical visualization of n-dimensional pareto front for decision-making in multiobjective optimization. *Information Sciences*, 178(20), 3908 – 3924.

Boada, Y., Reynoso-Meza, G., Picó, J., and Vignoni, A. (2016). Multi-objective optimization framework to obtain model-based guidelines for tuning biological synthetic devices: An adaptive network case. *BMC Systems Biology*. doi:10.1186/s12918-016-0269-0.

Bonilla-Petriciolet, A., Sharma, S., and Rangaiah, G.P. (2013). Phase equilibrium data reconciliation using multi-objective differential evolution with tabu list. In G.P. Rangaiah and A. Bonilla-Petriciolet (eds.), *Multi-objective optimization in chemical engineering: developments and applications*. John Wiley & Sons, 2013.

Buchler, N.E., Gerland, U., and Hwa, T. (2005). Nonlinear protein degradation and the function of genetic circuits. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27), 9559–9564.

Chellaboina, V., Bhat, S., Haddad, W., and Bernstein, D. (2009). Modeling and analysis of mass-action kinetics. *IEEE Control Systems Magazine*, 29(4), 60–78.

Chiang, A.W.T., Liu, W.C.C., Charusanti, P., and Hwang, M.J.J. (2014). Understanding system dynamics of an adaptive enzyme network from globally profiled kinetic parameters. *BMC Syst Biol*, 8, 4.

Egland, K.A. and Greenberg, E.P. (2000). Conversion of the Vibrio fischeri Transcriptional Activator, {LuxR}, to a Repressor. *Journal of Bacteriology*, 182(3), 805–811.

Gábor, A. and Banga, J.R. (2015). Robust and efficient parameter estimation in dynamic models of biological systems. *BMC Syst Biol*, 9, 74.

Kaltenbach, H.M.M., Dimopoulos, S., and Stelling, J. (2009). Systems analysis of cellular networks under uncertainty. *FEBS Lett*, 583(24), 3923–30.

Kim, J., Khetarpal, I., Sen, S., and Murray, R. (2014). Synthetic circuit for exact adaptation and fold-change detection. *Nucleic acids research*, 42(9), 6078–6089.

Leveau, J.H. and Lindow, S.E. (2001). Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria. *Journal of bacteriology*, 183(23), 6752–6762.

Lillacci, G. and Khammash, M. (2010). Parameter estimation and model selection in computational biology. *PLoS Comput Biol*, 6(3), e1000696.

Ma, W., Trusina, A., El-Samad, H., Lim, W.A., and Tang, C. (2009). Defining network topologies that can achieve biochemical adaptation. *Cell*, 138(4), 760–773.

Mendes, P. and Kell, D. (1998). Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, 14(10), 869–883.

Miettinen, K., Ruiz, F., and Wierzbicki, A.P. (2008). Introduction to multiobjective optimization: interactive approaches. In *Multiobjective Optimization*, 27–57. Springer.

Moles, C.G., Mendes, P., and Banga, J.R. (2003). Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome research*, 13(11), 2467–2474.

Picó, J., Vignoni, A., Picó-Marco, E., and Boada, Y. (2015). Modelling biochemical systems: from mass action kinetics to linear noise approximation. *Revista Iberoamericana de Automática e Informática Industrial RIAI*, 12(3), 241–252. doi:10.1016/j.riai.2015.06.001.

Prescott, T.P. and Papachristodoulou, A. (2014). Layered decomposition for the model order reduction of timescale separated biochemical reaction networks. *Journal of theoretical biology*, 356, 113–122.

Pronzato, L. and Pázman, A. (2013). Design of experiments in nonlinear models: Asymptotic normality, optimality criteria and small-sample properties. *Lecture Notes in Statistics*.

Reynoso-Meza, G., Blasco, X., Sanchis, J., and Martínez, M. (2014). Controller tuning using evolutionary multi-objective optimisation: current trends and applications. *Control Engineering Practice*, 28, 58–73.

Reynoso-Meza, G., Sanchis, J., Blasco, X., and Herrero, J.M. (2012). Multiobjective evolutionary algortihms for multivariable PI controller tuning. *Expert Systems with Applications*, 39, 7895 – 7907.

Reynoso-Meza, G., Sanchis, J., Blasco, X., and Martínez, M. (2010). Design of continuous controllers using a multiobjective differential evolution algorithm with spherical pruning. *Applications of Evolutionary Computation*, 532–541.

Samee, M.A.H., Lim, B., Samper, N., Lu, H., Rushlow, C., Jiménez, G., Shvartsman, S., and Sinha, S. (2015). A systematic ensemble approach to thermodynamic modeling of gene expression from sequence data. *Cell Systems*, 1(6), 396–407.

Srinivas, M. and Patnaik, L.M. (1994). Genetic algorithms: A survey. *Computer*, 27(6), 17–26.

Steuer, R., Gross, T., Selbig, J., and Blasius, B. (2006). Structural kinetic modeling of metabolic networks. *Proceedings of the National Academy of Sciences*, 103(32), 11868–11873.

Velasco-Carrau, J., Garca-Nieto, S., Salcedo, J.V., and Bishop, R.H. (2015). Multi-objective optimization for wind estimation and aircraft model identification. *Journal of Guidance, Control, and Dynamics*, 39(2), 372–389.

Villaverde, A.F., Bongard, S., Mauch, K., Müller, D., Balsa-Canto, E., Schmid, J., and Banga, J.R. (2015). A consensus approach for estimating the predictive accuracy of dynamic models in biology. *Comput Methods Programs Biomed*, 119(1), 17–28.