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_2CLab 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.

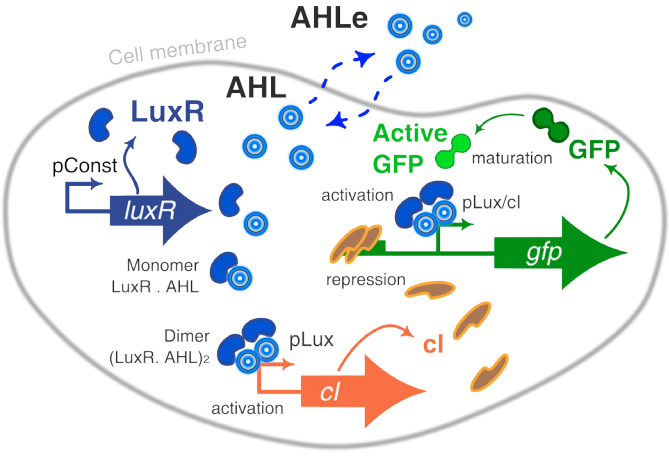

Fig. 1. Representation of a cell incorporating the engineered incoherent feed-forward loop synthetic circuit.

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-