Jugowice, 11th–15th September 2017

EVALUATION OF THE SYSTEM DESIGN METHOD OF ANALYSIS OF THE DYNAMICAL PROPERTIES OF NON-LINEAR SYSTEMS

Wojciech Bensz[1], Marek Kimmel[1,2]

[1]Silesian University of Technology
ul. Akademicka 16, 44-100 Gliwice, Poland
wojciech.bensz@polsl.pl

[2]Rice University
6100 Main Street, Houston, TX 77005, USA
kimmel@rice.edu

ABSTRACT

Qualitative methods of analysis of dynamic system properties possess a number of advantages over traditional methods based on quantitative models. Parameter estimation is particularly problematic in case of more complex or strongly non-linear models. System design space methodology allows for efficient identification and examination of qualitatively different system behaviours observed for varying values of parameters. The aim of presented work is to evaluate limitations of the design space method.

INTRODUCTION

Generalized Mass Action (GMA) system is one of the canonical representations of non-linear dynamical systems [6] which has the following form:

$$\begin{aligned}
\frac{dX_1}{dt} &= \sum_{k=1}^{P_1} \alpha_{1k} \prod_{j=1}^{n+m} X_j^{g_{1jk}} - \sum_{k=1}^{Q_1} \beta_{1k} \prod_{j=1}^{n+m} X_j^{h_{1jk}} \\
&\vdots \\
\frac{dX_{n_t}}{dt} &= \sum_{k=1}^{P_{n_t}} \alpha_{n_t k} \prod_{j=1}^{n+m} X_j^{g_{n_t jk}} - \sum_{k=1}^{Q_{n_t}} \beta_{n_t k} \prod_{j=1}^{n+m} X_j^{h_{n_t jk}} \\
0 &= \sum_{k=1}^{P_{n_t+1}} \alpha_{(n_t+1)k} \prod_{j=1}^{n+m} X_j^{g_{(n_t+1)jk}} - \sum_{k=1}^{Q_{n_t+1}} \beta_{(n_t+1)k} \prod_{j=1}^{n+m} X_j^{h_{(n_t+1)jk}} \\
&\vdots \\
0 &= \sum_{k=1}^{P_n} \alpha_{nk} \prod_{j=1}^{n+m} X_j^{g_{njk}} - \sum_{k=1}^{Q_n} \beta_{nk} \prod_{j=1}^{n+m} X_j^{h_{njk}}
\end{aligned} \tag{1}$$

where n_t is the number of dynamic variables; n_c, the number of auxiliary variables; $n = n_t + n_c$, the number of dependent variables; m, the number of independent variables; α_{ik}, the rate constant for the k-th positive term of the i-th equation; β_{ik}, the rate constant of the k-th negative term of the i-th equation; P_i and Q_i, the number of positive and negative terms of the i-th equation, respectively; g_{ijk} and h_{ijk}, the kinetic order of the influence of the j-th variable on the k-th positive and negative

term of the *i*-th equation, respectively; and X_j, the *j*-th variable. The first n_t variables are the dynamic variables, the next n_c are the auxiliary variables and the last m are the independent variables [3].

A sub-class of the GMA systems, called the S-systems, comprises of systems described by ordinary differential equations having only one positive and one negative term [8]. In the 1980s, S-systems were proposed as the simplest canonical form of non-linear systems exhibiting saturable and synergistic properties [8]. S-systems can be transformed into log-linear systems, facilitating the analysis of their dynamic properties. Efficient numerical integration algorithms were developed for S-systems that allowed a significant improvement in terms of computation times [8]. Methods of calculation of explicit symbolic solutions for non-zero steady states and of symbolic determination of local stability were also proposed [8]. With rapidly growing capabilities of the computers, the efficiency of numerical integration algorithms tailored for S-systems could no longer outweigh the limitations of S-system based approach in the biological systems modelling. However, the S-system concept has been used together with the more general GMA canonical form in a new methodology of qualitative dynamics analysis developed by Savageau and co-workers [3]. To understand this methodology, it is first necessary to introduce the concept of the design space of a biochemical system and other auxiliary definitions. The design of a biochemical system represented as a GMA ordinary differential equation system consists of the form and the kinetic orders of reactions described by these equations, i.e. the assumed exponents of the powers of biochemical species concentrations in the right hand sides [6]. In order to characterize all possible qualitative behaviours of a system assuming a given design, a wide range of parameter values needs to be taken into consideration.

Sampling the parameter space is an inefficient and computationally expensive way of searching for distinct qualitative behaviour. The system design space approach is based on two observations. First, at any moment of the system evolution, regardless of assumed parameter values and initial conditions, a maximal positive and a minimal negative term can be identified for each of the GMA system equations. Second, the terms of greatest magnitude of each sign influence the evolution of the system to the largest extent at that particular moment of time. Conditions for the dominance of an arbitrarily chosen terms of the *i*-th equation take form of inequalities [3]:

$$\begin{aligned} &\alpha_{ip_i} \prod_{j=1}^{n+m} X_j^{g_{ijp_i}} > \alpha_{ik} \prod_{j=1}^{n+m} X_j^{g_{ijk}} \forall k = \{1, 2, 3, \ldots, P_i | k \neq p_i\} \\ &\beta_{iq_i} \prod_{j=1}^{n+m} X_j^{h_{ijq_i}} > \beta_{ik} \prod_{j=1}^{n+m} X_j^{h_{ijk}} \forall k = \{1, 2, 3, \ldots, P_i | k \neq q_i\} \end{aligned} \quad (2)$$

Logarithmic transformation leads to a system of linear inequalities [3]:

$$\begin{aligned} &\log \alpha_{ip_i} + \sum_{j=1}^{n+m} g_{ijp_i} \log X_j > \log \alpha_{ik} + \sum_{j=1}^{n+m} g_{ijk} \log X_j \\ &\forall k = \{1, 2, 3, \ldots, P_i | k \neq p_i\} \\ &\log \beta_{iq_i} + \sum_{j=1}^{n+m} h_{ijq_i} \log X_j > \log \beta_{ik} + \sum_{j=1}^{n+m} h_{ijk} \log X_j \\ &\forall k = \{1, 2, 3, \ldots, P_i | k \neq q_i\} \end{aligned} \quad (3)$$

Dominance of single terms in each of the GMA system equations requires conditions in the form of a system of inequalities to be satisfied. By neglecting all but the dominant terms of all equations, we obtain an approximation of the GMA system in the form of a S-system. A set of dominance conditions in the form of a system of inequalities represents a region of dominance of an S-system in the variable and parameter space.

In order to characterize different qualitative behaviour of the system, the Design Space method assumes every possible combination of dominant positive and negative terms in every equation, yielding a pool of S-systems dominating in different regions of the combined parameter and variable spaces. Individual combinations of dominance assumptions are referred to as cases.

Steady states of the S-systems corresponding to the cases can be determined using linear algebra and represented as functions of the parameters of the original GMA system. By substituting steady state solutions for the state variables in a system of inequalities representing dominance conditions of an S-system, the regions of dominance are projected onto parameter space, creating the design space of the GMA system.

The dominant S-systems are nonlinear subsystems, presumed to approximate behaviour of the full system within defined regions of the parameter space. The aim of the design space method is to identify the so-called qualitative phenotypes, i.e. dynamical systems sharing the same design, but exhibiting qualitatively different dynamics due to differences in the parameter values. Qualitative phenotypes are analogous to phenotypes as defined in genetics, in the sense that they are special cases of a more general system that can exhibit various behaviour depending on its parameters, corresponding to different environmental conditions and genetic differences between individual organisms of the same kind [10]. Despite the broad range of possible applications of design space method in systems analysis, there are some caveats. First, the underlying assumption that S-systems can provide accurate approximations of the qualitative behaviour of more complex non-linear systems cannot be justified by a formal mathematical proof. Moreover, little is known concerning the coverage of the parameter space by the valid regions of phenotypic dominance (referred to as coverage further in this text) and its dependence on the complexity of the system.

The aim of this study is to examine the coverage utilizing the Design Space Toolbox V2 [3] and three models of differing complexity as examples. Secondary aim of the study is to evaluate the applicability of the design space method to qualitative analysis of a model that was shown before to undergo different kinds of bifurcations, dependent on a single bifurcation parameter.

MATERIALS AND METHODS

Design Space Toolbox V2 (DSTv2) C library [3] has been compiled from source codes on one of the nodes of the Ziemowit HPC cluster under the CentOS 7 operating system. IPython and Design Space Toolbox Python interface have been installed on the top of Python 2.7 on the same computer system. We chose the model of the p53-Mdm2 regulatory module [1] to perform the tests of the method implemented in the Design Space Toolbox v2. The original model consists of seven ordinary differential equations, of which six describe p53 and Mdm2 protein concentrations in different forms and one describes DNA damage. Additional algebraic equation describes the p53 degradation ratio, dependent on the DNA damage, denoted k_{d2}. In the original paper authors performed bifurcation analysis, in which the k_{d2} was treated as a bifurcation parameter, hence neglecting its dependence on the DNA damage [1]. A Saddle-Node, Saddle-Node-Loop and Hopf bifurcations were found when changing values of the k_{d2} parameter, resulting in a complex dynamics.

For the purpose of evaluation of the Design Space Toolbox performance, we similarly treated k_{d2} as a parameter and omitted the DNA damage variable, as only k_{d2} was dependent on its value in the original model. Remaining six ordinary differential equations were recast into generalized mass action (GMA) system canonical form [6] in order to obtain a system of equations compatible with the DSTv2 input format:

$$
\begin{aligned}
\frac{p53_t}{dt} &= ks_{53} - kd_{53p}p53_t - kd_{53}p53_{uu} \\
\frac{p53_u}{dt} &= k_f M_{nuc} p53_t + k_r p53_{uu} - 2k_f M_{nuc} p53_u - k_f M_{nuc} p53_{uu} - k_r p53_u - kd_{53p}p53_u \\
\frac{p53_{uu}}{dt} &= k_f M_{nuc} p53_u - k_r p53_{uu} - kd_{53p}p53_{uu} - kd_{53}p53_{uu} \\
\frac{M_{nuc}}{dt} &= V_r k_i MP_{cyt} - V_r k_o M_{nuc} - kd_2 M_{nuc} \\
\frac{M_{cyt}}{dt} &= k_{s2p} + k_{s2} p53_t^3 X100^{-1} + k_{deph} MP_{cyt} - k_{d2p} M_{cyt} - k_{ph} M_{cyt} X200^{-1} \\
\frac{MP_{cyt}}{dt} &= k_{ph} M_{cyt} X200^{-1} + k_o M_{nuc} - k_{deph} MP_{cyt} - k_i MP_{cyt} - k_{d2p} MP_{cyt} \\
0 &= J_s^3 + p53_t^3 - X100 \\
0 &= J_{ph} + p53_t - X200
\end{aligned}
\tag{4}
$$

Left hand sides of the equations are derivatives of the dependent variables with respect to time. X100 and X200 are auxiliary variables introduced to yield a canonical GMA form of the model, equal to denominators of fractions found in the original equations.

The first part of the DSTv2 analysis we performed, consisted of following steps:

1. Definition of the design space for the model in GMA form;
2. Validation of the phenotypes of all possible sub-systems found by the DSTv2 for the GMA system;
3. Evaluation of design space coverage by the dominance regions of the validated S-systems.

The parameters, unless specifically constrained to match *a priori* knowledge of the modelled system, can assume any positive real values, hence the parameter space extends from zero to positive infinity in all dimensions. Since all dominance regions of all system designs considered in this study are bounded polytopes, the actual greatest values of the parameters belonging to sums of their respective dominance regions are finite numbers, facilitating coverage evaluation by means of Monte Carlo method. Evaluation of design space coverage was performed by Monte Carlo simulation, where parameter values were randomly and independently chosen from the range between zero and an assumed maximal value of the parameter. In order to choose reasonable assumptions as to the maximal parameter values, we have performed following steps. First we have used DSTv2 built-in method to find a parameter set lying on one of the vertices of every dominance polytope. Further, we have searched those parameter sets for the single greatest coordinate in each dimension separately.

Each vector of random parameters generated in the course of Monte Carlo simulation was checked against dominance conditions of all valid S-systems. Fraction of vectors lying within at least one region of dominance was calculated as an estimate of the spatial coverage of design space by the dominance regions found by DSTv2.

The same procedure for parameter space coverage determination was performed for two simpler models, both used before to demonstrate the possibilities of design space methodology [10]. First model consists of two differential equations with four dependent variables (including two auxiliary variables) and nine independent variables. Its design space contains 16 valid phenotypes [3]:

$$
\begin{aligned}
\frac{X_1}{dt} &= \alpha_1 X_4^{-1} + \alpha_1 \rho_1 X_2^2 K_1^{-2} X_4^{-1} - \beta_1 X_1 \\
\frac{X_2}{dt} &= \alpha_2 X_1 X_5^{-1} + \alpha_2 X_1 \rho_2 X_3^2 K_2^{-2} X_5^{-1} - \beta_2 X_2 \\
0 &= 1 + X_2^2 K_1^{-2} - X_4 \\
0 &= 1 + X_3^2 K_2^{-2} - X_5
\end{aligned}
\tag{5}
$$

Second model, called Relaxation Oscillator [9] is more complex yet it has only 15 valid phenotypes:

$$\begin{aligned}
\frac{X_1}{dt} &= \alpha_1 \rho_1^{-1} X_5^{-1} + \alpha_1 K_{2A}^{-2} X_2^2 X_5^{-1} + \alpha_1 \rho_1^{-1} K_4^{-2} X_4^2 X_5^{-1} - \beta_1 X_1 \\
\frac{X_2}{dt} &= \alpha_2 X_1 - \beta_2 X_2 \\
\frac{X_3}{dt} &= \alpha_3 \rho_3^{-1} X_6^{-1} + \alpha_3 K_{2R}^{-2} X_2^2 X_6^{-1} - \beta_3 X_3 \\
\frac{X_4}{dt} &= \alpha_4 X_3 - \beta_4 X_4 \\
0 &= 1 + K_{2A}^{-2} X_2^2 + K_4^{-2} X_4^2 - X_5 \\
0 &= 1 + K_{2R}^{-2} X_2^2 - X_6
\end{aligned} \tag{6}$$

We couldn't reproduce the original bifurcation analysis of the p53-Mdm2 model utilizing design space methodology proposed in [10] since the case encompassing original parameter set was not recognized as a valid case by DSTv2. Instead, we have used DSTv2 to calculate steady state solution of the dominating S-system and find a representative parameter set for each valid phenotype in order to perform numerical simulations to inspect the qualitative behaviour. We have performed numerical simulations of the p53-Mdm2 model in unchanged form with Matlab ODE15s solver, switching the irradiation value to 1 after 30 hours of equilibration for 30 minutes and back to 0 for another 40 hours. The results of numerical simulations were searched for sustained oscillations after irradiation exposure, which we were expecting to obtain.

RESULTS

1518 valid cases were found in the design space constructed for the p53-Mdm2 model. 10 000 iterations of the Monte Carlo simulation were performed. 1047 parameter vectors were found to lay within the dominance boundaries of at least one qualitative phenotype, yielding approximately 10% coverage of the parameter space. As expected, parameter space coverage by dominance regions of the p53-Mdm2 system design was significantly scarcer than for the designs of simpler models presented in [10]. The coverage of the parameter space was complete or nearly complete for both, the simple design with 16 valid phenotypes and more complex relaxation-oscillator design. 100% of parameter sets generated for the former design and 98% for the latter fell inside valid dominance regions of their respective phenotypes.

Counting the total number of encompassing phenotype dominance regions for every random parameter set generated, allowed us to calculate the fraction of parameter space occupied by overlapping parts of dominance polytopes. We found no overlap between the valid dominance regions of the p53-Mdm2 design, i.e. all the random points belong to either none or exactly one region. The overlapping parts of dominance regions of the simplest design occupy approximately 0.005% of the total volume of parameter space examined. The relaxation-oscillator design produces the largest overlap in dominance regions, where approximately 68% of random space parameter points belong to a single dominance polytope, 28% belong to three dominance polytopes, 12% belong to 5 dominance polytopes and no overlaps of 2 or 4 regions were found.

Neither the total coverage nor the fraction of overlapping volumes of dominance polytopes seems to be directly correlated with the number of dependent or independent variables in the GMA system. The simplest design has two dependent variables, the relaxation-oscillator has 4, and the p53-Mdm2 has 6. The number of dependent and auxiliary variables is: 9, 15 and 18 respectively. The significant differences in examined geometric features of the phenotypic dominance regions suggest, that the interconnectivity of the components of the modelled system influence its design space properties to a much greater extent than the sole number of variables in the system.

It was not possible to directly compare the results of the design space based qualitative analysis of the p53-Mdm2 model with bifurcation analysis [1], because case encompassing original parameter set was not validated due to lack of closed form steady state of the corresponding S-system. Out of 1518 valid qualitative phenotypes found by DSTv2 there was only one exhibiting damped oscillations of the total p53 level. Moreover the amplitude of oscillations is far too small and the mean value is rather high compared to simulations results shown in [1].

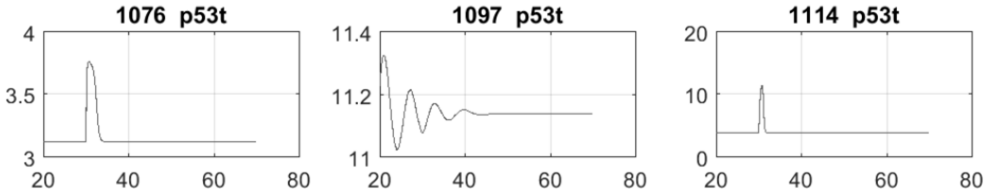

Figure 1. Examples of simulation results obtained with parameters and steady state solutions used as initial conditions generated with DSTv2. Total p53 quantity (dimensionless) plots are shown, time is expressed in hours, numbers above plots are case designations.

In order to rule out the possibility that p53-Mdm2 model is highly sensitive to parameter changes we have performed simple analysis where the parameters of the model were changed randomly, one at a time by up to 30% and while retaining similar qualitative behaviour. After the irradiation a digital oscillatory response was observed. Changes of parameter values resulted in varying number of pulses and times between them, but not to a qualitative change of the system response.

CONCLUSIONS

We have shown that the coverage of the parameter space by the dominance regions of qualitatively distinct phenotypes of a dynamical system may vary at least between a 10% and 100% of its total volume. More sophisticated methods are needed in order to allow a better understanding of the underlying geometry of the dominance regions. Based on anecdotal examples, in smaller systems, coverage by S-systems is close to 100%, while the rather small uncovered regions, correspond to systems behaving “badly”, for example “exploding” in finite times (Michael Savageau, personal communication). Based on discussion further on, this is not the case for the p53 model.

One explanation for a significant fraction of the parameter space remaining hollow for more complex designs, such as design based on the p53-Mdm2 model, is that some of the dominance regions could be narrow along one or a number of the dimensions, but elongated along the remaining dimensions extending the total parameter space taken into consideration. The reasoning behind this hypothesis is that most of the parameters do not influence the magnitude of most terms of such GMA systems.

The lower coverage indicates greater average distances between the boundaries of the dominance regions, as we only considered a finite subsets of the infinite parameter spaces, encompassing all the dominance regions discovered by the DSTv2 for respective designs. Conversely, greater distance between dominance regions boundaries could show the advantage of design space methodology over any method based on sampling the parameter space in terms of its efficiency. On the other hand, representative parameter sets, generated by DSTv2, corresponding to different qualitative phenotypes of the p53-Mdm2 model did not allow us to observe expected qualitative properties. Lack of overlapping regions suggests that p53-Mdm2 model is not capable of exhibiting some complex behaviours, such as bi-stability [3], [10], which is not the case [1]. Large number of discovered dominance regions further complicates the analysis and their sparse distribution over the parameter space means that most of the regions lie beyond the physically reasonable boundaries of parameter values associated with biochemical reaction rates. All of this

clearly shows that the applicability of the design space method to system design of higher dimensionality may be limited.

ACKNOWLEDGMENTS

The research presented here was funded by Silesian University Technology grant BKM/506/RAU1/2016/16 and the Polish National Centre for Science granted by decision number: DEC-2012/05/B/NZ2/01618.

References:

[1] A. Ciliberto, B. Novak, and J. Tyson. *Steady States and Oscillations in the p53/Mdm2 Network.* Cell Cycle (Georgetown, Tex.) 4, no. 3 (2005), 488–93.

[2] E.O. Voit and M.A. Savageau. *Accuracy of Alternative Representations for Integrated Biochemical Systems.* Biochemistry 26, no. 21 (1987), 6869–80.

[3] J. Lomnitz and M.A. Savageau. *Design Space Toolbox V2: Automated Software Enabling a Novel Phenotype-Centric Modeling Strategy for Natural and Synthetic Biological Systems.* Frontiers in Genetics **7**, 2016.

[4] J. Lomnitz and M.A. Savageau. *Elucidating the Genotype–phenotype Map by Automatic Enumeration and Analysis of the Phenotypic Repertoire.* NPJ Systems Biology and Applications **1,** 2015.

[5] M.A. Savageau, *Biochemical Systems Analysis.* Journal of Theoretical Biology 25, no. 3 (1969), 370–79.

[6] M.A. Savageau, *Design Principles for Elementary Gene Circuits: Elements, Methods, and Examples.* Chaos: An Interdisciplinary Journal of Nonlinear Science 11, no. 1 (2001), 142–159.

[7] M.A. Savageau, *Optimal Design of Feedback Control by Inhibition.* Journal of Molecular Evolution 5, no. 3 (1975), 199–222.

[8] M.A. Savageau and E.O. Voit. *Recasting Nonlinear Differential Equations as S-Systems: A Canonical Nonlinear Form.* Mathematical Biosciences 87, no. 1 (1987), 83–115.

[9] M.R. Atkinson, M.A. Savageau, J.T. Myers and A.J. Ninfa *Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli.* Cell **113(5)** (2003), 597–607.

[10] R.A. Fasani and M.A. Savageau. *Automated Construction and Analysis of the Design Space for Biochemical Systems.* Bioinformatics 26, no. 20 (2010), 2601–9.