

# PSExplorer: whole parameter space exploration for molecular signaling pathway dynamics

Thai Quang Tung and Doheon Lee\*

<sup>1</sup>Department of Bio and Brain Engineering, KAIST 373–1 Guseong-dong, Yuseong-gu, Daejeon 305–701, Republic of Korea

Associate Editor: Trey Ideker

## ABSTRACT

**Motivation:** Mathematical models of biological systems often have a large number of parameters whose combinational variations can yield distinct qualitative behaviors. Since it is intractable to examine all possible combinations of parameters for non-trivial biological pathways, it is required to have a systematic strategy to explore the parameter space in a computational way so that dynamic behaviors of a given pathway are estimated.

**Results:** We present PSExplorer, a computational tool for exploring qualitative behaviors and key parameters of molecular signaling pathways. Utilizing the Latin hypercube sampling and a clustering technique in a recursive paradigm, the software enables users to explore the whole parameter space of the models to search for robust qualitative behaviors. The parameter space is partitioned into sub-regions according to behavioral differences. Sub-regions showing robust behaviors can be identified for further analyses. The partitioning result presents a tree structure from which individual and combinational effects of parameters on model behaviors can be assessed and key factors of the models are readily identified.

**Availability:** The software, tutorial manual and test models are available for download at the following address: <http://gto.kaist.ac.kr/~psexplorer>

**Contact:** tqtung@kaist.ac.kr; tqtung@gmail.com

Received on April 26, 2010; revised on July 4, 2010; accepted on July 25, 2010

## 1 INTRODUCTION

Computational modeling is a valuable tool to understand the complex biological systems. Such models often possess complex topological structures with multiple positive and/or negative feedbacks and a large number of parameters so that combinational perturbations on these parameter values can greatly affect qualitative behaviors of the models. Two important issues in this approach are to understand the scope of behaviors that the model can experience, and to identify key parameters regulating the behaviors.

Bifurcation analysis (Borisuk and Tyson, 1998) is one of well-developed methods for qualitative behavior analysis of dynamical models. Its application to systems biology is limited, however, since it can examine the effects of only a few parameters, mostly one or two at a time. Although a tool for the global parametric perturbation analysis of biochemical models is available (Zi *et al.*, 2008), it is focusing on parameter sensitivity analysis, i.e. a quantification of

individual effects of parameter variations respecting to quantitative changes in the model outputs.

In this note, we outline PSExplorer, an efficient computational tool to explore high-dimensional parameter space of biochemical models for identifying qualitative behaviors and key parameters. The software supports input models in SBML format. It provides a friendly graphical user interface allowing users to vary model parameters and perform time-course simulations at ease. Various graphical plotting features helps users analyze the model dynamics conveniently. Its output is a tree structure that encapsulates the parameter space partitioning results in a form that is easy to visualize and provide users with additional information about important parameters and sub-regions with robust behaviors.

## 2 METHODS

An ordinary differential equation (ODE) model of a biochemical network is often represented in the following form:

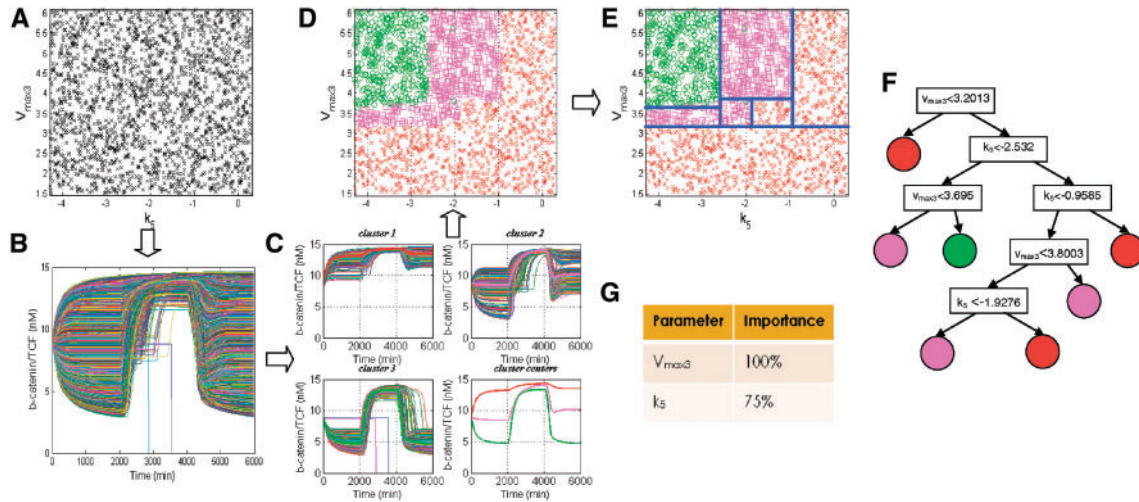
$$\frac{d\mathbf{X}}{dt} = f(\mathbf{X}, \boldsymbol{\theta}) ; \mathbf{X}(t=0) = \mathbf{X}_0 \quad (1)$$

where  $\mathbf{X}$  is a vector of state variables and  $\mathbf{X}_0$  a vector of their corresponding initial values;  $\boldsymbol{\theta}$  is a vector of kinetic parameters.

To begin with, users are required to specify a set of  $k$  parameters  $P_1, P_2, \dots, P_k$  ( $P_i$  can be a kinetic parameter or an initial value,  $i = 1, 2, \dots, k$ ) and their admissible value ranges to be perturbed; a set of target variables to represent the model behaviors; for convenience, let us say that there is only one target variable  $Y$ . The general framework of PSExplorer has six steps that are performed in a recursive manner. To help readers grasp main concepts, an illustrative example is provided in Figure 1 complementing to following brief explanations of the steps:

- (1) Latin hypercube sampling technique (Helton *et al.*, 2005) is used to generate  $N$  points in the parameter space (Fig. 1A);  $N$  is initially specified by users. Each sampling point represents a parameter combination  $\mathbf{p}^{(i)} = [p_{i1}, p_{i2}, \dots, p_{ik}]$  whose element  $p_{ij}$  is a randomly generated value of the parameter  $P_j$  within its admissible range.
- (2) For each  $\mathbf{p}^{(i)}$ , the software solves the set of ODEs in (1) to estimate temporal profile  $\mathbf{y}^{(i)}$  of the target variable, (Fig. 1B).
- (3) The simulated temporal profiles are grouped into  $c$  clusters (Fig. 1C) using a feature-based  $k$ -mean clustering algorithm. Each cluster is regarded as a qualitative behavior of the model. The software supports a set of features from which users need to select a subset that fits to their analysis purpose. The number of clusters is specified by users.
- (4) This step is to estimate a vector  $\mathbf{b}^{(i)} = [b_{i1}, b_{i2}, \dots, b_{ic}]$  for each  $\mathbf{y}^{(i)}$ ; where  $b_{ij}$  ( $j = 1, 2, \dots, c$ ) is the similarity measure between  $\mathbf{y}^{(i)}$  and  $j$ -th cluster center. For visualizing, we assign each  $\mathbf{p}^{(i)}$  to a cluster  $j$  if  $b_{ij}$  is the maximum value of  $\mathbf{b}^{(i)}$ , then points in each cluster is plotted with a distinct color in the parameter space as shown in Figure 1D.

\*To whom correspondence should be addressed.



**Fig. 1.** An example illustrating a 2-parameter perturbation analysis of a pathway model. (A) Parameter space is filled with sampling points. (B) Plotting of simulation results. (C) Clustering shows three distinctive clusters representing model qualitative behaviors. (D) Sampling points is colored respecting to clustering results. (E) The parameter space is nicely partitioned into sub-regions with different behaviors. (F) A tree represents the parameter space partitioning. (G) Parameter ranking estimated from the tree.

- (5) Let  $B_j(j=1,2,\dots,c)$  be a variable representing similarity measure between the  $j$ -th cluster center and simulated temporal profiles. Considering a regression problem where  $P_1, P_2, \dots, P_k$  are predictors and  $B_1, B_2, \dots, B_c$  are responses, a multi-response regression (Segal, 1992) is constructed using  $N$  learning cases  $\langle \mathbf{p}^{(i)}, \mathbf{b}^{(i)} \rangle (i=1,2,\dots,N)$ . Loosely speaking, the goal of this step is to partition the parameter space into sub-regions containing homogenous points (as shown in Fig. 1E). The partitioning is represented in a binary tree (Fig. 1F) whose terminal nodes correspond to the sub-regions.
- (6) Heterogeneous sub-regions are identified and additionally sampled with  $N_0$  (specified by users) points. These sampling points are proportionally distributed to the sub-regions respecting to their heterogeneities. The Steps 2–6 are then repeated iteratively until the total number of generated parameter combinations exceeding a maximum value. Note that in Step 2, simulations are performed only on a newly generated data and results are combined with the data of previous iterations.  $N$  is then replaced by  $N+N_0$  afterward.

Finally, PSExplorer presents the result tree from Step 5 of the final iteration in a graphical format where users can intuitively pinpoint regions of robust behaviors for further analyses. From the tree structure, the software evaluates parameters in terms of their importance and key parameters can be readily to be identified (Fig. 1G).

### 3 EXAMPLE

An application example of how to use PSExplorer to study system dynamics of computational models is provided in the user's manual of the software. The Wnt–ERK crosstalk model (Kim *et al.*, 2007) is selected as an example model. Fifty-five kinetic parameters of the model were varied from  $10^{-1}$  to  $10^1$  times of their reference values. PSExplorer has identified five distinct qualitative behaviors. Three among them were mentioned in a previous study explaining co-operatively working modes of the two pathways. In contrast, the two remaining ones show that ERK and Wnt can behave independently regardless the existence of a hidden positive feedback between them. Literature information has been found to support the existence of

such behaviors. We refer readers to the user's manual for a detail description of the example.

### 4 DISCUSSION

It is imperative to employ an *in silico* analysis to systematically identify the specific perturbations that have significant effects on the system behaviors, especially when numerous experiments on living systems are not practical. Globally perturbing model parameters to analyze system dynamics was used in previous studies (Hua *et al.*, 2006; Sung *et al.*, 2009). For automating the approach, PSExplorer has been developed from the framework proposed by Hua *et al.* (2006) with necessary refinements to boost computational performance.

Instead of using an exhaustive sampling approach, PSExplorer utilizes a novel recursive stratified sampling approach that can help explore high-dimensional spaces efficiently. On each recursion step, the Latin hypercube sampling requires less samples to be generated, thanks to its balancing distribution property. Then, the coupling of a clustering algorithm with a regression tree model provides a mean for a parameter space stratification, which allows the sampling process to concentrate on highly heterogeneous regions while ignoring homogeneous ones. As a result, the number of samples required to properly explore the parameter spaces can significantly reduce.

Last but not the least, one unique feature of the software that makes it more interesting than the sensitivity analysis technique is the ability to identify combinatory effects of parameter variations on model behaviors. The result tree not only identifies sensitive parameters but it also can tell how they combine to produce a specific behavior of interest.

**Funding:** WCU (World Class University) program (R32-2008-000-10218-0) and Korean Systems Biology Research Project (20100002164) of the Ministry of Education, Science and

Technology (MEST) through the National Research Foundation of Korea.

*Conflict of Interest:* none declared.

## REFERENCES

- Borisuk,M.T. and Tyson,J.J. (1998) Bifurcation analysis of a model of mitotic control in frog eggs. *J. Theor. Biol.*, **195**, 69–85.
- Helton,J.C. *et al.* (2005) A comparison of uncertainty and sensitivity analysis results obtained with random and Latin hypercube sampling. *Reliab. Eng. Syst Saf.*, **89**, 305–330.
- Hua,F. *et al.* (2006) Integrated mechanistic and data-driven modelling for multivariate analysis of signalling pathways. *J. R. Soc. Interface*, **3**, 515–526.
- Kim,D. *et al.* (2007) A hidden oncogenic positive feedback loop caused by crosstalk between Wnt and ERK pathways. *Oncogene*, **26**, 4571–4579.
- Segal,M.R. (1992) Tree-structured methods for longitudinal data. *J. Am. Stat. Assoc.*, **87**, 407–418.
- Sung,M.H. *et al.* (2009) Sustained oscillations of NF-kappaB produce distinct genome scanning and gene expression profiles. *PLoS One*, **4**, e7163.
- Zi,Z. *et al.* (2008) SBML-SAT: a systems biology markup language (SBML) based sensitivity analysis tool. *BMC Bioinformatics*, **9**, 342.