Multi-Criteria Optimization Based Experimental Design for Parameter Estimation of a Double Feedback Gene Switching Model

Vaibhav Maheshwari, Manoj Kandpal, Lakshminarayanan Samavedham*

*Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore
(*e-mail: chels@nus.edu.sg)*

Abstract: Despite the rapid increase in quantity and quality of experimental data in many fields of engineering and science, quantitative measurements of many cellular components are still relatively scarce. This work deals with estimating the parameters of a double feedback gene-switching model. To achieve the goal, a model-based design of experiment (MBDOE) approach for parameter estimation is employed. To overcome the problem of convergence in parameter estimation step (due to correlation among the parameters), a non-dominated sorting genetic algorithm (NSGA-II) based, multi-objective optimization (MOO) based MBDOE has been used. The parameter estimates obtained through the MOO based DOE as well as a standard alphabetical DOE technique are then compared with the known true values from the literature to highlight the efficacy of the MOO-MBDOE technique.

Keywords: Design of Experiments, Feedback Loop, Parameter Estimation, Gene Switching model, Multi-Criteria Optimization

1. INTRODUCTION

Every cellular system is composed of various macro/micro-molecules that are essential for its proper functioning. Various inter-connections among such molecules are often mapped in the form of a complex network which may represent different information-processing functions. Depending upon the type of information flow considered, these networks can be differentiated into genetic networks, signal transduction pathways, or metabolic networks. Within a complex network, some recurring sub-networks of particular function are designated as network motifs. Similar to engineering systems that are often built of recurring circuit modules, these biological network motifs are observed to obey similar principles to carry out key functions (Shen-Orr et al., 2002). Among such motifs, it has been found out that numerous biological systems comprise positive feedback as a key regulatory motif. These loops can have single, double, or multiple positive feedback regulatory effect. Some of the extensively studied processes are dynamics of a motif consisting of interlinked fast and slow positive feedback loops, which regulate for e.g., polarization of budding yeast, calcium signalling, Xenopus oocyte maturation, etc. (Brandman et al., 2005).

Description of such key regulatory motifs in the form of mathematical model can provide essential understanding of the complete network and pinpoint the importance of various parameters involved in modelling (Kim et al., 2012). Various quantitative, mechanistic modelling approaches, in particular, differential equation models have been applied to study such network motifs. These dynamic models furnish a detailed molecular and quantitative understanding of cellular information flow for such systems (Barkai and Leibler,

1997). Eventually, formulation of predictive models for such networks can help in studying complex diseases as well. In order to develop mechanistic models as *in silico* proxies for biological functions, researchers need to estimate the values of important parameters employed in such models. Prediction from these mathematical models essentially depends on the correct estimation of the unknown model parameters, which require collection of samples from *in vivo* or *in vitro* experiments (Michailidis, 2012).

An intelligent experiment design often provides better point estimates for unknown model parameters with limited resources. The work here is focused on the model based design of experiments (MBDOE), and subsequent parameter estimation of a network motif model. A detailed explanation of the model is provided in section 2. Traditional experimental designs, namely, D-, A-, and E-optimal designs are the most widely used forms of DOE techniques. These alphabetical designs provide the optimal moves for manipulated variable(s) and the optimal sampling instances for data collection. However, these designs often increase the correlation among unknown model parameters which can result in poor precision and poor point estimates of the unknown model parameters; thus violating the inherent objective of MBDOE techniques (Franceschini and Macchietto, 2008a). To overcome this drawback inherent in existing MBDOE techniques, a multi-objective optimization (MOO) based framework for MBDOE is employed (Maheshwari et al., 2013). The relevant details and drawbacks of traditional alphabetical designs, existing solutions, and the MOO based solution are detailed in section 3. In section 4, the parameter estimation results obtained from the alphabetical design and the MOO based MBDOE framework are compared – the results indicate that the MOO

based MBDOE framework outperforms the traditional alphabetical designs. In this work, only D-optimal design and its MOO counterpart are considered. The parameter estimates and their precision were obtained via Monte Carlo simulations. Finally, in section 5, conclusions and recommendations for future work are provided.

2. NETWORK MOTIF MODEL

As mentioned in section 1, a motif model is a recurring entity in a network. These network motif models often occur with single and dual-positive/negative feedback loops. In the dual positive feedback loops, two variables mutually activate a third variable. Furthermore, the two activating variables could be activated by an external stimulus. The schematic of such a network motif is presented in Figure 1. Corresponding set of model equations for the network motif is represented as equation 1.
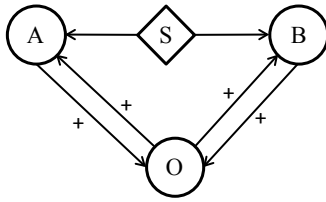

Figure 1: Schematic representation of dual feedback network motif. A and B mutually affect O, while S is external stimulus affecting A and B alone.

$$\frac{dA}{dt} = \tau_a \left(k_{min} - A + \frac{O^n}{O^n + ec_{50}}(1-A)S \right)$$

$$\frac{dB}{dt} = \tau_b \left(k_{min} - B + \frac{O^n}{O^n + ec_{50}}(1-B)S \right) \quad (1)$$

$$\frac{dO}{dt} = k_{on}^{out}(A+B)(1-O) - k_{off}^{out}O + k_{min}^{out}$$

The model presented here is normalized and non-dimensionalized, i.e. O, A, and B are concentrations between 0 and 1 (Brandman et al., 2005). Here, O is mutually activated by A and B, with a nonlinear Hill function (equation 2), which describes the relationship between the concentration of O and the rate of production of A and B.

$$h(O) = \frac{O^n}{O^n + ec_{50}} \quad (2)$$

where, n = 3 is the Hill coefficient describing cooperative binding and ec_{50} is the concentration for half-maximum response for the feedback. The activation of O is dependent on activation of A and B, which in turn can be activated by external stimulus S. Here, the external stimulus S simultaneously affect both A and B. The dynamics of motif when one of the inner loops is suppressed is also of interest. In such scenario, the system of equations reduces to two equations for A and O alone (assuming B is suppressed). In this work, we have considered the comprehensive motif i.e. when both inner loops are activated.

Like any mathematical model, the present model also comprises of a number of model parameters whose estimation is required before the motif can be plugged into the complete network model. The objective here is to design an experiment which, if implemented on the system, will provide information rich data for estimating the unknown model parameters. In this simulation work, the parameter estimates culminating from the experiment are compared with the known true values to adjudge the efficacy of MBDOE technique. The model parameters with their known true values are listed in Table 1 (Kim et al., 2012).

Table 1: Model parameters with their known values

Parameter	True parameter value
ec_{50}	0.35
k_{min}	0.01
k_{min}^{out}	0.001
k_{on}^{out}	2
k_{off}^{out}	0.3
τ_a	0.5
τ_a	0.008

3. MBDOE TECHNIQUES - DRAWBACK AND SOLUTION

A designed experiment is superior to intuitive experimental design for accurately estimating the parameters of dynamic systems with minimum resources (Bandara et al., 2009). MBDOE techniques are statistical procedures to select the best experimental settings corresponding to maximum information under pre-defined operational and budget constraints. As the name suggests, a mathematical model structure is the foremost requirement for MBDOE techniques. Additionally, a good initial guess for unknown model parameters is also required. The *handle* available with the experimenter is the number of samples, sampling instances, initial conditions for system states, permissible moves and corresponding switch time for external input, etc. Concisely, given the model structure and initial guess for its parameters, the aim of the MBDOE approaches is to suggest experimental designs that help us achieve a certain objective – one objective can be to minimize the parameter variance i.e. to make the elements of parameter variance-covariance matrix (V) small. Conventional experiment design thus involves minimizing a metric of V, or maximizing that of its inverse, the Fisher Information Matrix (FIM), which plays key role in MBDOE techniques for improvement of parameter precision. The MBDOE techniques culminate in an optimization problem with objective,

$$\Delta = min\,arg(V) \equiv max\,arg(FIM) \quad (3)$$

The FIM is ($p{\times}p$) symmetric matrix and defined as,

$$FIM(\hat{\theta}, \phi) = \sum_{i=1}^{n_y} \sum_{j=1}^{n_y} \sigma_{ij} G_i^T G_j \quad (4)$$

where, σ_{ij} is the $(i,j)^{th}$ element of inverse of variance-covariance matrix of experimental measurements, n_y is the number of measured states, $FIM(\hat{\theta}, \phi)$ is the information matrix with $\hat{\theta}$ as parameter guess and ϕ as vector of decision variables (here experimental conditions). $\mathbf{G_i}$ is s×p matrix of first order sensitivity coefficients for i^{th} measured state for n_{sp} sampling instances,

$$G_i = \left[\frac{\partial y_{is}}{\partial \theta_k}\right]; s = 1, 2, ..., n_{sp}; k = 1, 2, ..., p \quad (5)$$

Number of real-valued functions is used to quantify the FIM or V into a scalar metric. The most common criteria include:

1. *D-optimality criterion* maximizes the determinant of FIM, or equivalently, minimizes the determinant of the matrix V.
2. *A-optimality criterion* maximizes the trace of FIM, or equivalently, minimizes the trace of matrix V.
3. *E-optimality criterion* aims to maximize the smallest eigenvalue of FIM, or equivalently, minimizes the largest eigenvalue of matrix V.

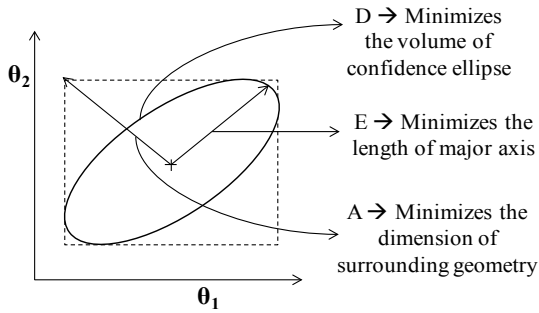The geometrical interpretation of these criteria is provided in Figure 2.

Figure 2: Geometrical interpretation of the D-, A-, and E-optimal design criteria for the case of two parameters.

As discussed, the objective of MBDOE techniques is to suggest the optimal input moves (perturbations) and optimal time point for sampling the outputs which will result in maximally informative data for estimating unknown model parameters with high degree of precision. However, precision should not be the only measure for good quality estimates. The de-correlation among model parameters is also important, as high correlation may plausibly lead to poor point estimates and/or poor precision of model parameters (Rodriguez-Fernandez et al., 2006). Correlation among parameters can also lead to convergence issues in the parameter estimation step. Due to correlation, change in one parameter can be offset by changes in others, resulting in a situation where various parameter combinations result in comparable values of the objective function. This can create problems for optimization algorithms because no definite direction can be found in which the objective function value will improve. This makes it difficult to obtain the unique estimates of model parameters, or even leads to inaccurate estimates for parameters. Interestingly, maximizing the

traditional MBDOE objective alone also increases the correlation among model parameters (Franceschini and Macchietto, 2008b). To overcome this issue, a multi-criteria optimization based MBDOE framework is proposed (Maheshwari et al., 2013). Before we discuss about MOO framework, it is important to understand the scenarios wherein correlation can manifest.

Correlation among parameters can be due to model structure (structural identifiability issue) and/or due to experimental data (practical identifiability issue). Both identifiability aspects can be understood from the simple algebraic models.

$$y_1 = k_1 k_2 x$$
$$y_2 = k_1 (1 - e^{-k_2 x}) \quad (6)$$

where, k_1, k_2, and x are positive. In the first model, estimation of the parameter combination $k_1 k_2$ is possible from experimental data, but individual estimation of k_1 and k_2 is impossible even with infinite amount of samples. Hence, the model is structurally unidentifiable. In the second model, it seems that both the parameters can be estimated. However, for smaller '$k_2 x$' values, $e^{-k_2 x} \simeq 1 - k_2 x$ and model 2 reduces to model 1, which is already known to be structurally unidentifiable. Thus, to estimate both the parameters in model 2, samples should be collected for larger values of 'x'. This makes model 2 an example where parameter correlation can be due to poor experimental data. The issue can, however, be dealt with using intelligent experimental design. It may be relatively easier to detect the practical/structural correlation for algebraic systems, but such insights are difficult to obtain for dynamic systems.

To address this correlation problem in dynamic systems, a number of solutions have been proposed; examples of such work include minimizing the correlation measure (Pritchard and Bacon, 1978) or anti-correlation designs (Franceschini and Macchietto, 2008b). Here, we have considered a different approach where both information and correlation criteria are accounted in an optimization framework. This culminates in an MOO framework where the fist criterion is the conventional objective of maximizing an information measure, while the second criterion corresponds to minimization of correlation among model parameters. The correlation measure can be calculated using variance-covariance matrix itself. The expression for r_{ij}^{th} element of correlation matrix is,

$$r_{ij} = \frac{V_{ij}}{\sqrt{V_{ii} V_{jj}}} \quad (7)$$

The final correlation matrix R and the objective function to be minimized ($\|R\|$) in MOO based MBDOE framework is,

$$R = \begin{bmatrix} 1 & r_{12} & . & r_{1n} \\ r_{21} & 1 & . & r_{2n} \\ . & . & . & . \\ r_{n1} & r_{n2} & . & 1 \end{bmatrix} \Rightarrow \|R\| = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}^2}; i > j \quad (8)$$

Among the D-, A-, and E-optimality criteria, the D-optimality criterion is popular and has been used previously for elucidating the parameter estimates for similar biological system (Bandara et al., 2009). Hence, we have studied the performance of D-optimal design only. The resulting MOO design is referred as DMOO design. Corresponding conflicting criteria are,

$$\text{obj1}: \ max \ |\text{FIM}|$$
$$\text{obj2}: \ min \ \|\text{R}\| \tag{9}$$

4. METHOD AND RESULTS

The definition of experimental design space is an important aspect in DOE. The design space is decided based on available experimental resources, such as what can be measured, how the manipulated variable should be perturbed, how many samples can be collected, etc. Here, the following design constraints are assumed.

1. A total of 15 samples can be collected.
2. Data is measurable for only state '*O*'.
3. Consecutive samples should have a minimum time interval of 5 time units. The maximum allowed time interval between consecutive samples is 200 time units. The last sampling instance decides the duration of experiment; however, the maximum experiment duration is restricted to 1200 time units.
4. The external stimulus is available handle for experimenter to perturb the system to elicit the most informative data. It is a binary variable which can either be 0 or 1, denoting switch off or switch on scenario, respectively.
5. It is assumed that total 5 control moves are allowed during the whole experiment duration. Also the minimum and maximum time between two control moves is 5 and 200 time units, respectively.

The mentioned design constrains results in a design space with 2 constraints and 24 decision variables, namely, 5 moves for external stimulus, 15 sampling instances, and 4 switching instances for manipulated variables. The resulting MOO problem is solved in MATLAB using NGPM (NSGA-II program in MATLAB) (Lin, 2012). As mentioned in section 3, the MBDOE framework requires the initial guess for unknown model parameters. In this simulation study, we assumed the initial parameter guess at 50% positive deviation from true parameter values. The true parameter values are presented in Table 1. The MOO problem results in non-dominated solutions, which are represented in the form of Pareto-optimal front. Here, each non-dominated solution corresponds to an optimal experimental design. The resulting front for DMOO design is presented in Figure 3, wherein the extreme point of Pareto-optimal front with maximum information-maximum correlation corresponds to D-optimal design. It can be observed from Figure 3 that as information maximizes, the correlation among model parameters also maximizes. The resulting external stimulus profile is shown in Figure 3.

In real experiment settings, once the optimal experimental design is obtained, it is used to generate data from the

experimental system. Here, the experimental set-up is the model itself with its true parameter values. To account for the noise characteristics of real system, the simulated data is corrupted by 10% relative noise. Resulting noisy data is used for estimation of "unknown" model parameters. The parameter estimates and corresponding precision were obtained via Monte Carlo simulations. A total of 50 parameter estimation runs were performed. The mean of 50 sets of parameter estimates is designated as obtained parameter estimates for optimal design, while standard deviation of 50 runs is used to calculate the parameter precision. The obtained estimates and corresponding parameter precision are presented in Table 2.
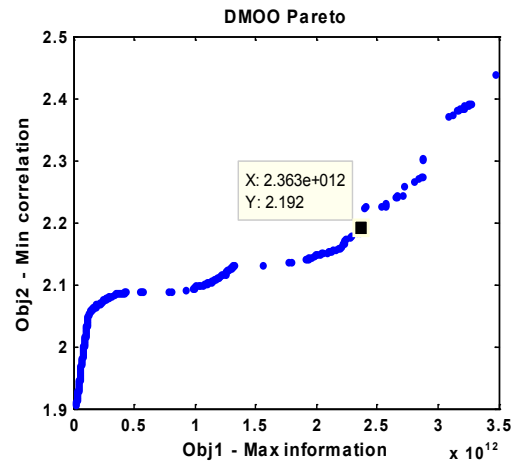

Figure 3: Pareto-optimal front for DMOO design. The selected point corresponds to chosen trade-off for comparison with D-optimal design
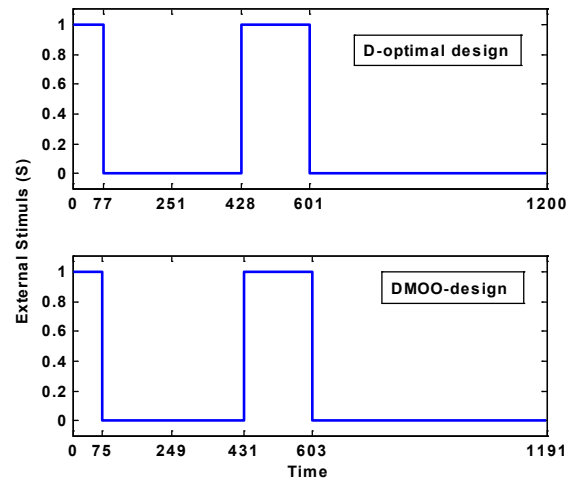

Figure 4: External stimulus profile for D- and DMOO design

In Figure 4, it can be observed that both D- and DMOO design results in visibly similar stimulus profile which will result in similar output profile for measured state '*O*'. However, the parameter estimates for DMOO design are better than that obtained in D-optimal design. The conclusion is based on normalized Euclidean distance measure (δ) from true parameter estimates (equation 10), which quantifies the error in estimation.

$$\delta = \sqrt{\sum_{i=1}^{n} \left(\frac{\theta_{i,true} - \theta_i}{\theta_{i,true}} \right)^2} \times 100 \qquad (10)$$

Table 2: Point estimates and 95% confidence interval with D-optimal and DMOO design

θ	D-optimal design			DMOO design		
ec_{50}	0.41	±	0.22	0.40	±	0.22
k_{min}	0.0094	±	0.0046	0.0094	±	0.0042
k_{min}^{out}	0.0009	±	0.0012*	0.001	±	0.0013*
k_{on}^{out}	2.94	±	1.45	2.85	±	1.58
k_{off}^{out}	0.40	±	0.19	0.38	±	0.16
τ_a	0.46	±	0.55*	0.47	±	0.58*
τ_a	0.0085	±	0.0025	0.0084	±	0.0023

* Statistically insignificant parameters

The smaller the δ, the closer are the obtained estimated parameter values to the true parameter values. The δ_D and δ_{DMOO} were 37% and 28%, respectively. Despite the similar external stimulus profile, the DMOO design results in better point estimates. This can be attributed to sampling instances, at which the data are sampled and subsequently used for parameter estimation. There is a subtle difference between sampling instances in both D- and DMOO designs. The difference is especially noticeable towards the completion of experiment (Table 3). Hence, for similar experimental efforts, the DMOO design results in improved point estimates for the gene network motif model.

Table 3: Sampling protocol for D-optimal and DMOO designs

Design protocol	Sampling instances (in mins)
D-optimal design	5, 13, 71, 76, 81, 233, 238, 342, 446, 451, 604, 755, 800, 1000, 1200
DMOO design	5, 10, 73, 78, 83, 235, 240, 345, 449, 454, 606, 742, 793, 993, 1191

In Table 2, parameters k_{min}^{out} and τ_a are statistically insignificant. To improve the precision of those parameters, either more samples should be collected or states other than 'O' should be measured.

5. CONCLUSIONS

A gene network motif model is a recurring model in global gene network. To estimate the parameters of such network motif, MBDOE techniques can play an important role. The drawback of existing experimental design techniques has been highlighted. To overcome the drawback an MOO based MBDOE framework is employed. The D-optimal design and its MOO based counterpart DMOO are compared for parameter estimates. For same experimental efforts, the

DMOO design results in better point estimates when compared with traditional D-optimal design. The performance of other alphabetical designs and their MOO based counterpart will be assessed in future.

REFERENCES

BANDARA, S., SCHLÖDER, J. P., EILS, R., BOCK, H. G. & MEYER, T. 2009. Optimal experimental design for parameter estimation of a cell signaling model. *PLoS computational biology,* 5**,** e1000558.

BARKAI, N. & LEIBLER, S. 1997. Robustness in simple biochemical networks. *Nature,* 387**,** 913-917.

BRANDMAN, O., FERRELL, J. E., JR., LI, R. & MEYER, T. 2005. Interlinked Fast and Slow Positive Feedback Loops Drive Reliable Cell Decisions. *Science,* 310**,** 496-498.

FRANCESCHINI, G. & MACCHIETTO, S. 2008a. Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science,* 63**,** 4846-4872.

FRANCESCHINI, G. & MACCHIETTO, S. 2008b. Novel anticorrelation criteria for model-based experiment design: Theory and formulations. *AIChE Journal,* 54**,** 1009-1024.

KIM, K.-K. K., CHEONG, K. S., CHEN, K. & BRAATZ, R. D. Parameter estimation, analysis, and design of synthetic gene switching models: System behavior- and performance-based approaches. IFAC ADCHEM, 2012. 946-951.

LIN, S. 2012. *NGPM -- A NSGA-II Program in Matlab v1.4* [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/31166 [Accessed Dec 2012 2012].

MAHESHWARI, V., RANGAIAH, G. P. & SAMAVEDHAM, L. 2013. Multiobjective Framework for Model-based Design of Experiments to Improve Parameter Precision and Minimize Parameter Correlation. *Industrial & Engineering Chemistry Research,* 52**,** 8289-8304.

MICHAILIDIS, G. 2012. Statistical Challenges in Biological Networks. *Journal of Computational and Graphical Statistics,* 21**,** 840-855.

PRITCHARD, D. J. & BACON, D. W. 1978. Prospects for reducing correlations among parameter estimates in kinetic models. *Chemical Engineering Science,* 33**,** 1539-1543.

RODRIGUEZ-FERNANDEZ, M., MENDES, P. & BANGA, J. R. 2006. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *Biosystems,* 83**,** 248-265.

SHEN-ORR, S., MILO, R., MANGAN, S. & ALON, U. 2002. Network motifs in the transcriptional regulation network of Escherichia coli. *Nature Genetics,* 31**,** 64-68.