# Specifying Informative Experiment Stimulation Conditions for Resolving Dynamical Uncertainty in Biological Systems

Thembi Mdluli[1], Michael Pargett[1], Gregery T. Buzzard[2], Ann E. Rundell[1]

***Abstract*—A computationally efficient model-based design of experiments (MBDOE) strategy is developed to plan an optimal experiment by specifying the experimental stimulation magnitudes and measurement points. The strategy is extended from previous work which optimized the experimental design over a space of measurable species and time points. We include system inputs (stimulation conditions) in the experiment design search to investigate if the addition of perturbations enhances the ability of the MBDOE method to resolve uncertainties in system dynamics. The MBDOE problem is made computationally tractable by using a sparse-grid approximation of the model output dynamics, pre-specifying the time points at which the input or experimental perturbations can be applied, and creating scenario trees to explore the endogenous uncertainty. Consecutive scenario trees are used to determine the best input magnitudes and select the optimal associated measurement species and time points. We demonstrate the effectiveness of this strategy on a T-Cell Receptor (TCR) signaling pathway model.**

## I. Introduction

Model-based design of experiments (MBDOE) techniques pose optimization problems to determine the experimental design that will collect maximally informative data. Major approaches towards MBDOE are concentrated towards 1) reducing parameter uncertainty [1], [2], [3], 2) discriminating between mathematical models [2], [4], [5], and 3) resolving dynamical uncertainty [6], [7]. Adoption of MBDOE integrates predictions made by a mathematical model within the experimental design process to avoid doing experiments that are potentially costly, time consuming and insufficiently informative. The MBDOE strategy is particularly informative for uncertain biological systems that can exhibit a broad range of dynamical responses.

A majority of the current MBDOE methods rely on local optimal designs that choose an experiment which maximizes an objective related to the Fisher Information Matrix (FIM) [8]. The local FIM-based methods are usually limited by the need for an initial estimate of the parameters and risk being stuck in local minima. Since biological system models are often nonlinear models and may possess many uncertain model parameters, the FIM-based approaches are not fully robust. Some global approaches have been proposed to overcome the challenges of FIM-based strategies. Most use Monte Carlo simulation to estimate confidence intervals to select maximally informative experiments [9], [10]. However, these methods are very computationally expensive, resulting in a need for a computationally tractable algorithm.

Computationally efficient global MBDOE methods focusing on reducing uncertainties in the model output dynamics to gain information about the system were proposed in [6], [7]. Their approaches reduced the computational burden of most global strategies by utilizing sparse-grid interpolation to approximate the model dynamics over the parameter space. They identified regions of the uncertain parameter space that resulted in similar (indistinguishable) model output dynamics, instead of identifying the true values of the parameters. Although quite efficient, these two MBDOE strategies focus on finding the best measurement points to resolve the model output dynamics, without considering possible stimulation (experimental inputs). Herein, we extend the global MBDOE approach presented in [6], [7] to a more holistic specification of experiments by introducing a method to determine the best experimental stimulation or perturbation levels in addition to the measurement points. The main contribution of this work is a computationally efficient method for performing *de novo* experiment design that differs from most existing work which tends to either select optimal experimental perturbations or measurements but not both. Previous MBDOE strategies would select the best experimental stimulations or perturbations from a predefined discrete subset of experimental factor choices. This may result in sub-optimal experiment designs [11]. Our approach performs *de novo* experiment design to determine the optimal stimulation magnitudes and measurements within a continuous feasible space rather than selecting them from predefined experiments.

## II. Methods

Our strategy considers a mathematical representation of the biological system in the form of non-linear ODEs,

$$\begin{aligned} \dot{x} &= f(x,u,\theta,t), \quad x(t_0) = x_0, \\ x_M &= C_M x, \quad x_T = C_T x, \end{aligned} \tag{1}$$

where $x \in \mathcal{X} \subset \mathbb{R}^n$ are the states, $\theta \in \Omega \subset \mathbb{R}^p$ are the model parameters, and $u \in \mathcal{U} \subset \mathbb{R}^{n_u}$ are the experimental inputs. In this work, the experimental input represents the experimental stimulation or perturbation which excites the biological system. The uncertainty in the model dynamics for

Research supported by National Science Foundation Career ECC 0846572 and DMS-0900277 grants.

[1] T. Mdluli (tmdluli@purdue.edu), M. Pargett and A. E. Rundell (rundell@purdue.edu) are affiliated with the Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN 47906.

[2] G. T. Buzzard is affiliated with the Department of Mathematics, Purdue University, West Lafayette, IN 47906.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

this deterministic model results from the uncertainty in the parameter values. A major challenge to modeling biological systems is that not all states are measurable but many of them are of interest in understanding the system behavior. We denote $x_M \in \mathbb{R}^{n_M}$ as the experimentally measurable states, and $x_T \in \mathbb{R}^{n_T}$ denotes a subset of the internal states of the system which are targeted for dynamical resolution. Measuring $x_M$ helps bound the estimated dynamics on the unmeasured states.

Our MBDOE strategy minimizes the uncertainty in the targeted model state dynamics, $x_T$ by specifying the experiment design in terms of an optimal input $u$ that perturbs the biological system and determines the associated maximally informative measurement points, $x_{M_i}(\tau_{M_i})$ , where $M_i$ indicates which of the measurable states should be measured at discrete time-points $\tau_{M_i}$.

Given an initial set of data, the first step of our MBDOE algorithm is to determine the dynamical behavior of the target states that are consistent with the available data. To do this, the uncertain parameter space is screened using sparse grids as in [6] to identify data consistent parameters. Herein, a parameter set is considered data consistent if it generates model output dynamics that fit the available data to within a standard deviation of each available data point,

$$\Omega_A = \{\theta : \hat{x}_M(t) - \sigma(t) \leq x_M(\theta, t) \leq \hat{x}_M(t) + \sigma(t)\}, \quad (2)$$

where $x_M(\theta, t)$ denotes the measured state values simulated by the model with parameter set $\theta$ at time $t$. $\hat{x}_M(t)$ and $\sigma(t)$ are the mean and standard deviation of the available experimental data, respectively. Note that $\Omega_A \subseteq \Omega$.

Since the available data for some biological systems tend to be sparse and noisy, there may be a large number of acceptable parameters. To reduce the computational burden of our MBDOE approach, a representative subset of the acceptable parameters is selected to cover all the experimentally distinguishable dynamics of the target states. Acceptable parameters are first clustered using a mean-distance clustering method so that parameters that give similar target state dynamics will be grouped together. This is similar to the approach used in [6], but here we restrict the dynamical clusters to the target state dynamics, and we sample equally from each of the dynamical clusters. These parameters form the representative parameter set, denoted by $\Omega_R$. Note that $\Omega_R \subseteq \Omega_A$.

The aim of the MBDOE strategy is to minimize the variability of the target state dynamics generated by the set of representative parameters. We use consecutive scenario trees (described below) to approximate a solution to the two-step objective optimization problem:

$$u^* = \arg\min_{u \in \mathcal{U}} \xi\left(x_T(u, \theta, t)\right), \quad (3)$$

$$(M_i^*, \tau_{M_i}^*) = \arg\max_{\{M_i, \tau_{M_i}\}} \xi\left(x_M(u^*, \theta, t)\right) \quad (4)$$

for $\theta \in \Omega_R$. We denote the triple $D^* \triangleq (u^*, M_i^*, \tau_{M_i}^*)$ as the optimal design. Herein, we define a distinguishability metric (DM), denoted by $\xi(x(u, \theta, t))$, to quantify the resolvable variability in the measured and target state dynamics generated by the representative parameters,

$$\xi\left(x(u, \theta, t)\right) = \max_{\theta \in \Omega_R} \frac{\sqrt{\text{var }\left(x(u, \theta, t)\right)}}{E\left(\sigma_x(t)\right)},$$

where,

$$\text{var }\left(x(u, \theta, t)\right) = \sum_{j=1}^{d} p(\theta_j)\left(x(u, \theta_j, t) - \mu_x(t)\right)^2$$

for $\theta_j \in \Omega_R \subset \mathbb{R}^{d \times p}$ is the weighted predicted dynamical variance of the state dynamics and,

$$E(\left(\sigma_x^2(t)\right) = \begin{cases} \zeta_b + \zeta_s |x_i(u, \theta, t)| + \zeta_t |\dot{x}_i(t)|, \ x_i(t) \in x_M \\ 1, \ x_i(t) \in x_T, \end{cases}$$

is the expected experimental variance for the measured states and unity for the target states. Here $x_i$ is the $i$th component of the state $x(t)$. For the predicted dynamical variance, the $\mu_x(t)$ is the mean of the states at time $t$ over all parameters in the representative parameter set and $p(\theta_j)$ are probabilities of the state dynamics according to how they fit the available data assuming a Gaussian noise distribution. For the experimental variance estimate, the $\zeta_b$ is the constant background variance, $\zeta_s$ is the linear contribution to the variance, and $\zeta_t$ is the variance contribution due to imprecise experimental time-sampling and resolution of the sparse grid approximation tool. The optimal measurement points are chosen as the points with the largest predicted DM in the measured states under the optimal input magnitudes that minimize the predicted DM of the target state dynamics.

For computational efficiency, this MBDOE optimization problem is solved using sparse-grid approximations of the target and measurable state dynamics over the uncertain parameter and input design space. This approach was used previously in [6], [7] over the uncertain parameter space and has been extended in this work to also consider the input design space.

To approximate the solution for the optimization problems, we use consecutive scenario trees to explore the endogenous uncertainty that arises from the unknown parameter space in a computationally efficient manner. We adapt the scenario tree process for parallel MBDOE from [7] to allow the input to be applied at different time points for each level of the tree. The number and time points for input administration are specified *a priori* while the magnitude of the inputs is determined by the MBDOE optimization problem. The first scenario tree is initialized by choosing the input from a LHS samples on the input design space to be applied at the first time point that minimizes the distinguishability metric of the target state dynamics. The best measurement point for that input will be the one with the largest distinguishability metric and give rise to the next three branches of the tree. The branches are constructed by predicting the measurement data to be in the 10th, 50th, and 90th percentiles of the dynamics according to their different probability weights. For each of the consecutive branches, the best input to be applied at the next time point is

optimized given the previous optimal input(s) are applied. The three branches give different optimal inputs designs due to the different probability weights of how the associated representative parameters fit each predicted data. Repeating this process, each node can stem up to three branches to form the next level in the tree. At this stage of the MBDOE algorithm, the number of levels of the scenario tree is defined by the number of time points at which the inputs can be administered. With

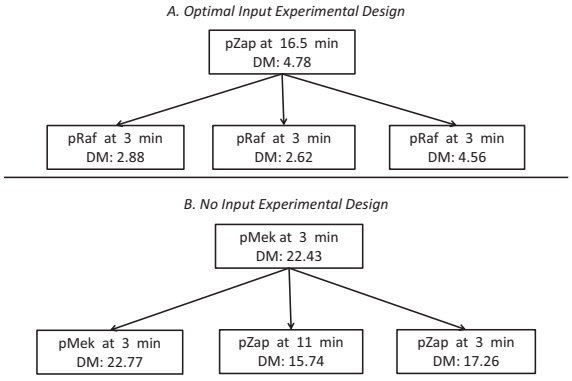


Fig. 1. Scenario tree designs comparing the optimized input experiment design. **A**. Against the unperturbed design. **B**. The optimized input sequence is 0.24 PMA and 0.16 U0126 at 0 minutes and 0.15 PMA and 0.05 U0126 at 15 minutes. The initial distinguishability metric (DM) value for the target states before taking any measurements is 22.83. The worst case final DM value given the optimal input experiment design is 4.56 which is better than the worst case DM value given by the unperturbed design, 22.77. Thus, the optimized input experiment design is more informative.

the addition of each potential measurement, the probabilities used in the MBDOE objective formula are updated following Bayes theorem with the likelihood of the data assumed to follow a Gaussian noise distribution:

$$p_n(\theta) = \frac{p(x|\theta)p_{n-1}(\theta)}{\tilde{p}(x)},$$

$$p(x|\theta) = \frac{1}{\sqrt{2\pi\,(E(\sigma_x^2))}} \exp\left(\frac{x(u,\theta,\tau) - \mu_x}{E(\sigma_x)}\right)^2,$$

$$\mu_x = E(x(u,\theta,\tau)) = \sum_{j=1}^{d} p(\theta_j)x(u,\theta_j,\tau).$$

The posterior parameter weight, $p_n(\theta)$ under the observation is given by the prior weight of the parameter $p_{n-1}(\theta)$ and the likelihood function $p(x|\theta)$ which is the probability of observing the predicted measurement data $x$ at sample time $\tau$ for the model parameterized by $\theta$. $\tilde{p}(x)$ is the normalization constant.

Each branch of the last level of the initial scenario tree defines a potential optimal input sequence from the top node to the bottom. A new scenario tree is created for each potential input sequence to determine the best measurement points and evaluate the DM value for the target states. The input design sequence, and its associated measurement points, corresponding to the minimum target state DM value is selected as the optimal experimental design.

## III. Exemplar Model and Results

The effectiveness of our strategy is demonstrated through a TCR intracellular pathway with a mathematical model proposed by in [13]. The model has 32 nonlinear ordinary differential equations (ODEs) and 53 reaction parameters. There are 3 measurable states: pZap, pRaf and pMek, and 4 target states; Zapb, pSHP1, pZap, and pErk. There are 2 experimental stimuli that could be applied: phorbol myristate acetate (PMA) and U0126. The input magnitudes are normalized between 0 and 1 and they can be administered at 0 minutes and again at 15 minutes. We allow a total of 4 measurements to be taken in the duration of the experiment. Therefore, the initial scenario tree has 2 levels and the second set of scenario trees will have 2 levels that specify up to 4 nodes that specify the measurement points. Initially, $\Omega_A$ contains 4097 parameters, from which we sample 100 representative parameters ($\Omega_R$) from computed dynamical clusters.

The final scenario tree for the experiment design is represented in Figure 1 for the optimized input experiment design in comparison with an unperturbed experiment design. The DM value is greatly reduced from the initial value of 22.77 to 4.78 by taking a single measurement of pZap at 16.5 min with the optimized input applied. In comparison, taking the first measurement on the unperturbed system does not add much information to resolve the target dynamics as shown in the scenario tree Figure 1B.

Additional measurements further reduce the variability of the target states output dynamics. To compare the information content of the two experiment designs shown in Figure 1, the biological system is simulated *in silico* for the optimized input sequence and the unperturbed cases to obtain four additional measurements with 10% Gaussian noise. The original acceptable parameter space $\Omega_A$ is screened for acceptable parameters that are consistent with the four new measurements for each of the two experiments designed creating parameter subsets $\Omega_{AOI} \subset \Omega_A$ for the optimized input experiment and $\Omega_{AM} \subset \Omega_A$ for the optimized measurement specification experiment (without additional perturbation). In Figure 2, the cyan (light blue) lines illustrate the range of the target state dynamics in accordance with the optimized input design. In Figure 3, the red lines indicate the reduced variability in the dynamics of the target states obtained with just optimized measurement specification. To illustrate the reduction in the uncertainty of the target state dynamics for both of these experiment designs, the initial range of the target output dynamics generated by the uncertainty in the acceptable parameter space is also provided in Figure 2 and 3 with the dark blue lines. A comparative study of the dynamics resolved with and without the optimized experimental inputs is presented in Figure 4. It is clear that our method outperforms the previous MBDOE method, as we observe the uncertainty in the target state dynamics are further reduced by the introduction of this added input optimization problem.

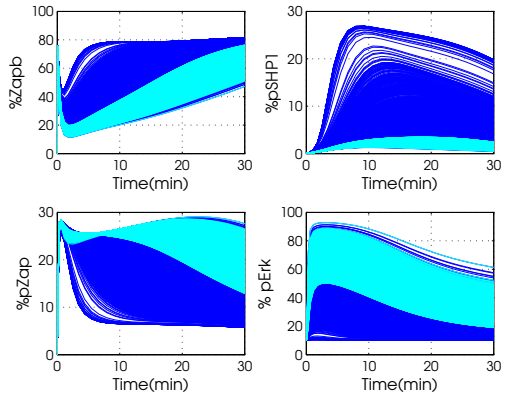


Fig. 2. Illustration of the reduction of the uncertainty in the target state dynamics for the unperturbed system after performing the optimized input experiment *in silico*. The dark blue lines indicate the initial range of dynamics simulated by all acceptable parameters, $\theta \in \Omega_A$. The cyan (light blue) lines show the target dynamics that are data consistent after the optimized input experiment, simulated with $\theta \in \Omega_{AOI}$.

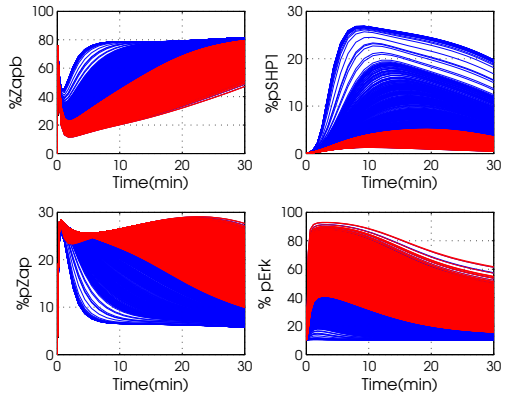


Fig. 3. Illustration of the reduction of the uncertainty in the target state dynamics for the unperturbed system after performing the optimized measurement specification experiment that does not alter the input *in silico*. The dark blue lines indicate the initial range of dynamics simulated by all $\theta \in \Omega_A$. The red lines show the target dynamics that are data consistent after optimized measurements were taken, simulated with $\theta \in \Omega_{AM}$.

## IV. Conclusions

In this paper, we proposed a computationally efficient and tractable MBDOE strategy that determines the experimental factors (both measurement points and stimulation levels) to minimize the target dynamic uncertainties. We illustrated the performance of the proposed method on a T-cell exemplar model *in silico* and compared our method to a previous MBDOE strategy that only specified the measurement points. We demonstrated the effectiveness of optimizing for experimental inputs in improving the MBDOE method for uncertainty resolution in the target dynamics. Our method has limitations as some assumptions and pre-specifications are required for tractability, which is the objective for future work.

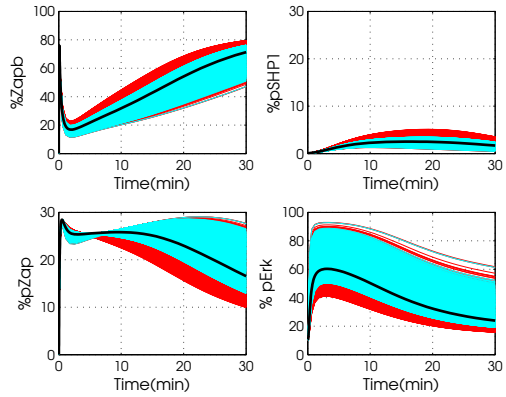


Fig. 4. Illustration of the enhancement achieved by optimizing experimental stimulation magnitudes in the MBDOE algorithm. The uncertainty in the target state dynamics is more reduced for the optimized experimental input design (cyan lines) compared to the previous method that only specified the measurements (red lines). The black line, denoting the actual target dynamics simulated with the nominal parameters, is within the uncertainty ranges refined by both MBDOE methods.

## References


[1] M. Bezeau and L. Endrenyi, *Design of experiments for the precise estimation of dose-response parameters: the Hill equation*, J. Theor. Biol., vol. 123, no. 4, pp. 415–430, Dec. 1986.

[2] G.E. Rovati, D. Rodbard, and P.J. Munson. *DESIGN: Computerized optimization of experimental design for estimating $K_d$ and $B_{max}$ in ligand binding experiments.*. Analytical biochemistry 184.1 (1990): 172–183.

[3] M. K. Transtrum and P. Qiu, *Optimal experiment selection for parameter estimation in biological differential equation models*, BMC Bioinformatics, vol. 13, p. 181, Jan. 2012.

[4] T. Maiwald, C. Kreutz, A. C. Pfeifer, S. Bohl, U. Klingmüller, and J. Timmer, *Dynamic pathway modeling: feasibility analysis and optimal experimental design.*, Ann. N. Y. Acad. Sci., vol. 1115, pp. 212–220, Dec. 2007.

[5] D. Skanda and D. Lebiedz, *A robust optimization approach to experimental design for model discrimination of dynamical systems*, Math. Program., vol. 141, no. 12, pp. 405–433, Mar. 2012.

[6] M. M. Donahue, G. T. Buzzard, and A.E. Rundell, *Experiment design through dynamical characterisation of non-linear systems biology models utilising sparse grids.*, IET Syst. Biol., vol. 4, no. 4, pp. 249–262, Jul. 2010.

[7] J. N. Bazil, G. T. Buzzard, and A. E. Rundell, *A global parallel model based design of experiments method to minimize model output uncertainty.*, Bull. Math. Biol., vol. 74, no. 3, pp. 688–716, Mar. 2012.

[8] G. Franceschini and S. Macchietto, *Model-based design of experiments for parameter precision: State of the art*, Chem. Eng. Sci., vol. 63, no. 19, pp. 4846–4872, Oct. 2008.

[9] K. Cho, *Experimental Design in Systems Biology , Based on Parameter Sensitivity Analysis Using a Monte Carlo Method: A Case Study for the TNF $\alpha$-Mediated NF- $\kappa$B Signal Transduction Pathway*, vol. 79, no. 11, 2003.

[10] X. Huan and Y. M. Marzouk, *Simulation-based optimal Bayesian experimental design for nonlinear systems*, J. Comput. Phys., vol. 232, no. 1, pp. 288–317, Jan. 2013.

[11] W. Dong, X. Tang, Y. Yu, R. Nilsen, R. Kim, J. Griffith, J. Arnold, and H.-B. Schttler, *Systems biology of the clock in Neurospora crassa.*, PLoS One, vol. 3, no. 8, p. e3105, Jan. 2008.

[12] P. Weber, A. Kramer, C. Dingler, and N. Radde, *Trajectory-oriented Bayesian experiment design versus Fisher A-optimal design: an in depth comparison study.*, Bioinformatics, vol. 28, no. 18, pp. 535–541, Sep. 2012.

[13] Y. Zheng, and A.E. Rundell. *Comparative study of parameter sensitivity analyses of the TCR-activated Erk-MAPK signaling pathway.* IEE Proceedings-Systems Biology 153.4 (2006): 201–211.