

RESEARCH ARTICLE

# Efficient Optimization of Stimuli for Model-Based Design of Experiments to Resolve Dynamical Uncertainty

Thembi Mdluli<sup>1\*</sup>, Gregory T. Buzzard<sup>2</sup>, Ann E. Rundell<sup>1\*</sup>

**1** Weldon School of Biomedical Engineering, Purdue University, West Lafayette, Indiana, United States of America, **2** Mathematics Department, Purdue University, West Lafayette, Indiana, United States of America

\* [tmluli@purdue.edu](mailto:tmluli@purdue.edu) (TM); [rundell@purdue.edu](mailto:rundell@purdue.edu) (AR)



**OPEN ACCESS**

**Citation:** Mdluli T, Buzzard GT, Rundell AE (2015) Efficient Optimization of Stimuli for Model-Based Design of Experiments to Resolve Dynamical Uncertainty. *PLoS Comput Biol* 11(9): e1004488. doi:10.1371/journal.pcbi.1004488

**Editor:** Sergei L. Kosakovsky Pond, University of California San Diego, UNITED STATES

**Received:** October 24, 2014

**Accepted:** August 5, 2015

**Published:** September 17, 2015

**Copyright:** © 2015 Mdluli et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This research was supported in part by the NSF grant DMS-0900277. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

This model-based design of experiments (MBDOE) method determines the input magnitudes of an experimental stimuli to apply and the associated measurements that should be taken to optimally constrain the uncertain dynamics of a biological system under study. The ideal global solution for this experiment design problem is generally computationally intractable because of parametric uncertainties in the mathematical model of the biological system. Others have addressed this issue by limiting the solution to a local estimate of the model parameters. Here we present an approach that is independent of the local parameter constraint. This approach is made computationally efficient and tractable by the use of: (1) sparse grid interpolation that approximates the biological system dynamics, (2) representative parameters that uniformly represent the data-consistent dynamical space, and (3) probability weights of the represented experimentally distinguishable dynamics. Our approach identifies data-consistent representative parameters using sparse grid interpolants, constructs the optimal input sequence from a greedy search, and defines the associated optimal measurements using a scenario tree. We explore the optimality of this MBDOE algorithm using a 3-dimensional Hes1 model and a 19-dimensional T-cell receptor model. The 19-dimensional T-cell model also demonstrates the MBDOE algorithm's scalability to higher dimensions. In both cases, the dynamical uncertainty region that bounds the trajectories of the target system states were reduced by as much as 86% and 99% respectively after completing the designed experiments *in silico*. Our results suggest that for resolving dynamical uncertainty, the ability to design an input sequence paired with its associated measurements is particularly important when limited by the number of measurements.

## Author Summary

Many mathematical models that have been developed for biological systems are limited because the complex systems are not well understood, the parameters are not known, and available data is limited and noisy. On the other hand, experiments to support model development are limited in terms of costs and time, feasible inputs and feasible

measurements. MBDOE combines the mathematical models with experiment design to strategically design optimal experiments to obtain data that will contribute to the understanding of the systems. Our approach extends current capabilities of existing MBDOE techniques to make them more useful for scientists to resolve the trajectories of the system under study. It identifies the optimal conditions for stimuli and measurements that yield the most information about the system given the practical limitations. Exploration of the input space is not a trivial extension to MBDOE methods used for determining optimal measurements due to the nonlinear nature of many biological system models. The exploration of the system dynamics elicited by different inputs requires a computationally efficient and tractable approach. Our approach plans optimal experiments to reduce dynamical uncertainty in the output of selected target states of the biological system.

## Introduction

Since experiments can be expensive and time consuming, it is important that they are planned to generate useful data. Traditional design of experiments is a well established field and has led to many advances in biology and medicine. The data obtained from strategically designed experiments has facilitated the creation of mathematical models that relate experimental stimuli to measurable outcomes. These models typically describe the system's input-output relationship but fail to capture or encode knowledge of the system's internal mechanisms and processes. Mechanistic and semi-mechanistic mathematical models encode the current understanding of the internal processes of the biological system even though many of these internal states or species are not directly measurable. These mechanistic models can be used to support optimal experiment design that considers the current knowledge of the system interactions and practical experimental constraints. In recent literature this type of experiment design has been referred to as model-based design of experiments (MBDOE). MBDOE produces experiments meant to reduce some measure of uncertainty in the associated model while respecting cost, time and resource constraints. Most MBDOE strategies can be categorized by three types of objectives: (1) reducing model parameter uncertainty [1–7], (2) discriminating among possible models [8–13], and (3) reducing dynamical uncertainty [14–17]. This work advances current abilities to design experiments to resolve the trajectories of target states of a biological system model, thereby reducing its dynamical uncertainty.

Many of the MBDOE strategies that support reduction of parameter uncertainty and model discrimination rely on linear approximations that are locally optimal to design an experiment by optimizing a criterion of the Fisher Information Matrix (FIM) [15, 18–21]. Such techniques use the local sensitivities of parameters to design an optimal experiment which requires an initial estimate of the unknown parameters. Most biological system models are not well characterized, as data is limited and noisy, so initial estimates of the model parameters are inaccurate. Furthermore, biological models are typically nonlinear, so a poor initial parameter estimate is likely to result in a local minimum, which may result in sub-optimal experiment design. Alternatively, the Sigma Point method refines estimates of the characteristic values of the parameter statistics [22, 23]. This method is computationally efficient and gives a better approximation of the covariance matrix when compared to the FIM-based and boot-strap methods. However, this method is not applicable when the biological system is not well characterized with existing experimental data. These local methods for MBDOE are limited to scenarios when initial estimates for the model parameters or their distributions are already fairly well known.

To overcome the shortcomings of local MBDOE strategies, global approaches that consider the entire uncertain parameter range have been developed. Global optimization has been used to improve the FIM-based design [24] and local deterministic optimization of the likelihood with multi-start initial parameter estimates [25] have been previously employed. Some Bayesian MBDOE methods use Monte Carlo generated estimates of the parameter confidence intervals to design a global optimal experiment [16, 26–30]. A comparison of the FIM-based and Bayesian-based MBDOE methods by Weber et al. [31] shows that Bayesian methods give more accurate and informative designs. However, all of these global methods are computationally prohibitive for large numbers of highly uncertain parameters and mathematical models that are computationally intensive. The limitation due to computation has been partially abrogated by methods that utilize surrogate models to approximate the mathematical model of the system. Herein, as in our previous work [14, 15], we use a sparse grid interpolation tool to approximate the biological model.

To find the optimal design a search has to be performed over all of the experimental factors that define the feasible experimental perturbations and potential measurements. A complete exploration of the full design space over the uncertain parameter space for an optimal experiment design is impractical due to the combinatorial explosion of possible experimental factors evaluated for all possible model parameterizations. To solve the problem, most MBDOE strategies constrain the design space to a subset of the possible experiments. This reduces the design space by assuming a small predefined set of possible measurement points, measurable species and/or input levels [9, 31–33]. Herein, we define the input to be the exogenous experimental factors, perturbations and/or stimuli that are applied to the biological system. Other designs have considered finding an optimal input sequence while the times and measurement species are defined *a priori* [2, 10, 24, 34]. To consider both optimal measurements and an optimal input sequence, some strategies [16, 35] evaluate the optimality of a pre-defined set of specified inputs and measurement points. All of these approaches may result in sub-optimal experiment designs if the best experiment is not among the specified experimental options.

Ideally, experiment design determines the optimal solution within a feasible space defined by all the possible experimental measurements and inputs. Since different inputs can elicit dramatically different dynamics for some nonlinear systems, evaluating experiments for optimal inputs is not necessarily a trivial extension of the global techniques for identifying optimal measurements. Herein, we describe and evaluate a computationally efficient and tractable method for performing global MBDOE to define optimal experimental inputs and measurements to reduce the uncertainty in continuous nonlinear system dynamics. The method utilizes the sparse grid surrogate model technique first employed for global experimental designs presented in [14, 15], which selected only optimal measurements and not optimal inputs. We proposed an earlier version [36] of our input design MBDOE algorithm that used a combination of scenario trees and sparse grid interpolation to screen the input design space for optimal inputs that will generate diverse dynamics of the uncertain nonlinear system. This paper is an extension of that work. Herein we utilize a greedy search to determine the input magnitude to be applied using a method that is similar to that in [37] which determined near-optimal measurements in polynomial time. Our MBDOE algorithm is made even more computationally efficient and robust than our previous sparse grid based strategies due to the uniform sampling over the dynamical space via careful selection of representative parameters, the optimization criteria for optimal experiment design for a target system, and the repeated update of probability weights over the dynamical space using predicted data. Our new MBDOE method develops a strategic experiment design that yields highly informative data that can be used to constrain the system dynamics to dynamical uncertainty regions.

In the methods section, we describe the experiment design problem and define the methods used to solve the problem as sequential optimizations to first select an optimal input sequence and, subsequently, specify the associated optimal measurement pairs (species and time points). In the results section, we explore the optimality of the derived experiment design solution using a small 3-dimensional Hes1 oscillatory model and demonstrate the scalability of this MBDOE strategy with a 19-dimensional T-cell receptor model. The global applicability, support of hybrid experiment designs, and limitations of this MBDOE approach are discussed to conclude the paper.

## Materials and Methods

### Dynamical Model Definition

This MBDOE algorithm is designed to resolve dynamical uncertainties associated with the biological system by reducing the predicted variance in the system dynamics. The system is described by nonlinear ordinary differential equations of the general form:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}, \boldsymbol{\theta}', t), \quad \mathbf{x}(t_0) = \mathbf{x}_0 \tag{1}$$

where  $\mathbf{f}$  is a deterministic smooth function of  $\mathbf{x} \in \mathbb{R}^{n_x}$ , the states of the system,  $\boldsymbol{\theta} \in \mathbb{R}^{n_p}$ , the known model parameters,  $\boldsymbol{\theta}' \in \mathbb{R}^{n_{p'}}$ , the unknown model parameters, and  $\mathbf{u} \in \mathbb{R}^{n_u}$  the exogenous inputs. Using control vector parameterization (CVP) [38], we represent  $\mathbf{u} = [\mathbf{u}(\tau_1), \dots, \mathbf{u}(\tau_N)] \in \mathbb{R}^{n_u \times N}$ , as vector-valued function of the system inputs that can change value at each discrete time point,  $\tau_j$ , where  $j = 1, \dots, N$ . There is uncertainty in the dynamics of the system which is generated by the uncertainty in the values of the unknown model parameters,  $\boldsymbol{\theta}' \in \Omega$  where  $\Omega$  is a compact set.

It is common in biological systems that not all states of this system are experimentally measurable. All feasible measurements for this system are modeled by:

$$\mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}, \boldsymbol{\theta}', t), \tag{2}$$

where  $\mathbf{h}$  is a smooth vector-valued function that relates to the system's internal states described in Eq (1) and  $\mathbf{y} \in \mathbb{R}^{n_m}$ . We abbreviate Eq (2) by the notation  $\mathbf{y}(\mathbf{u}, \boldsymbol{\theta}', t)$ . The user defines the number of measurements that they would like to have designed,  $K$ , as well as the set of possible times for measurement,  $\mathbb{T}$ , by either specifying discrete time points or defining a time resolution,  $\delta t$ , between  $T_I$  and  $T_F$ , the initial and final model simulation times, respectively.

Our MBDOE algorithm designs experiments that reduce dynamical uncertainty in a target system. This target system is a subset of the system states termed, target states,  $\mathbf{x}_T$ , when stimulated by a target input,  $\mathbf{u}_T$ . This approach employs the concept of a target system also used to evaluate an alternative MBDOE strategy [16].

We consider dynamical uncertainty in the measurable states,  $\mathbf{y}$ , and the target states,  $\mathbf{x}_T$ , generated by the uncertainty in the unknown parameters,  $\boldsymbol{\theta}'$ . As a result of limited and noisy data, often biological system models will fit the data with a wide range of parameters values and associated dynamics. Therefore, for an optimal design, we need to consider all the possible data-consistent dynamical scenarios which are a function of the parameter space,  $\Omega$ .

### Optimization Problem Definition

The optimal experiment,  $D^* \in \mathbb{D}$ , defines the optimal input stimulation and optimal measurement pairs that resolve the target state dynamics. This design is defined by a piecewise input sequence,  $\mathbf{u}^*$  (Eq (3)), and associated measurements  $M^* = \{(m_k, t_k): k = 1, 2, \dots, K\} \in \mathbb{M}$ , which defines  $K$  measurement pairs by the index of a measurement species  $m_k$  and its corresponding

measurement time  $t_k$ .

$$\mathbf{u}^* = \begin{cases} \mathbf{u}_1^* & \tau_1 \leq t \leq \tau_2 \\ \vdots & \vdots \\ \mathbf{u}_j^* & \tau_j \leq t \leq \tau_{j+1} \\ \vdots & \vdots \\ \mathbf{u}_N^* & \tau_N \leq t \leq T_F \end{cases} \quad (3)$$

We choose  $D^*$  as the experimental design that maximizes a measure of information gained from an experiment

$$D^* = \underset{\mathbf{u} \in \mathbb{R}^{n_u \times N}, M \in \mathbb{M}}{\operatorname{argmin}} \gamma(\mathbf{u}) \quad (4)$$

The target state dynamical uncertainty (TDU),  $\gamma(\mathbf{u})$ , is quantified by the sum of the maximum variance in each target state that results when analyzed assuming the input sequence,  $\mathbf{u}$ , has been applied and measurements,  $M$ , exist to constrain the uncertain parameter space:

$$\gamma(\mathbf{u}) = \sum_i \max(\operatorname{Var}(x_{T_i}(t)) | (\mathbf{u}, M)) \quad (5)$$

where the maximum variance of the  $i^{\text{th}}$  target state,  $\max(\operatorname{Var}(x_{T_i}(t)) | (\mathbf{u}, M))$ , is chosen across the simulated time  $t$ , from the initial,  $T_i$  to the final time,  $T_F$ .

The ideal solution for the above optimization problem simultaneously solves for the optimal input vector,  $\mathbf{u}^*$ , together with the optimal measurements,  $M^*$ . This is generally a computationally intractable problem due to the combinatorial explosion of all possible selections for the experiment design complicated by the expense of evaluating the design with the model for all possible uncertain parameter values. Our approach proposes a computationally efficient method to approximate both the optimal input vector and associated optimal measurement points by breaking the problem into smaller computationally feasible optimization problems. The first optimization problem solves for the input vector using a greedy method by minimizing the value of TDU at each iteration assuming a single best measurement is taken:

$$\mathbf{u}^* = \underset{u_j \in \mathbb{R}^{n_u}}{\operatorname{argmin}} \gamma(\mathbf{u}), j = 1, \dots, N \quad (6)$$

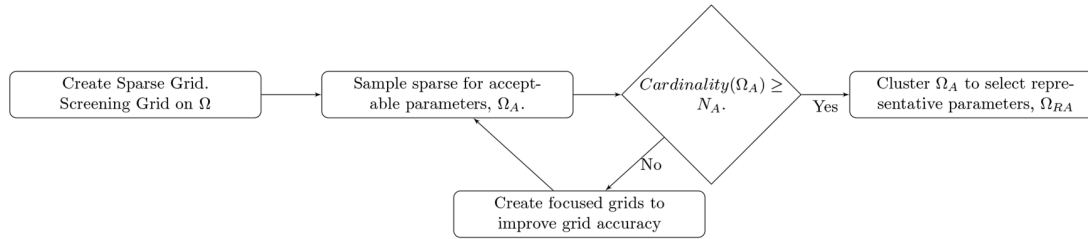
The optimal input vector is then used in the second optimization to determine the multiple optimal measurements points. This combined solution of input and measurements approximates the optimal experiment design defined in Eq (4).

A flow-chart of the MBDOE algorithm is given in Fig 1 to display the sequence of events in determining the optimal experiment design: (a) identify representative parameters that maintain the diversity of simulated dynamics that fit the existing data, (b) determine the optimal input vector that minimizes TDU, and (c) specify associated multiple measurement pairs given an optimal input vector.

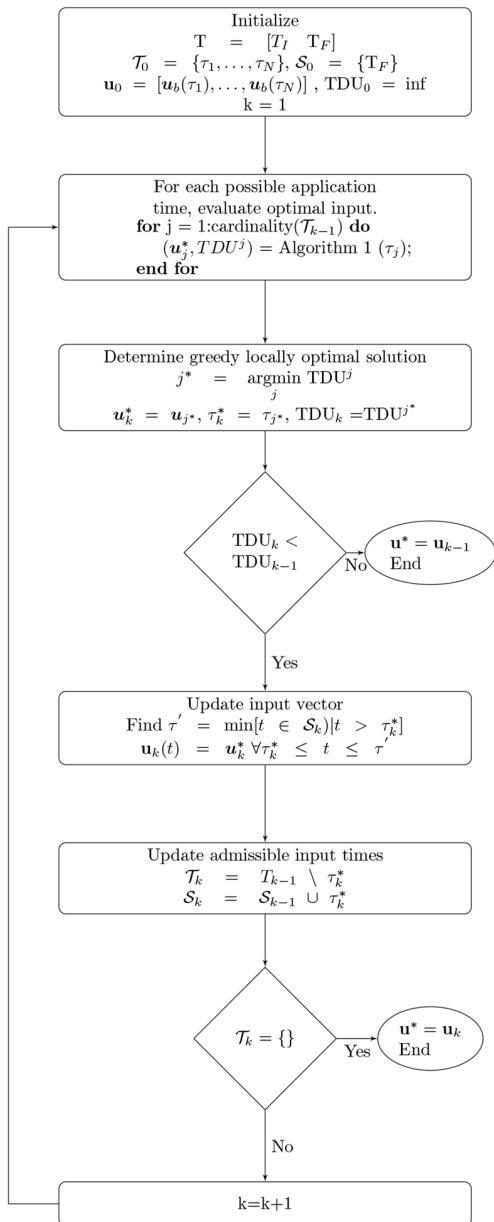
### Identify Representative Parameters

The first step of this process screens the uncertain parameter space,  $\Omega$ , to identify the space of acceptable parameters,  $\Omega_A$ . The acceptable parameters are those that support model

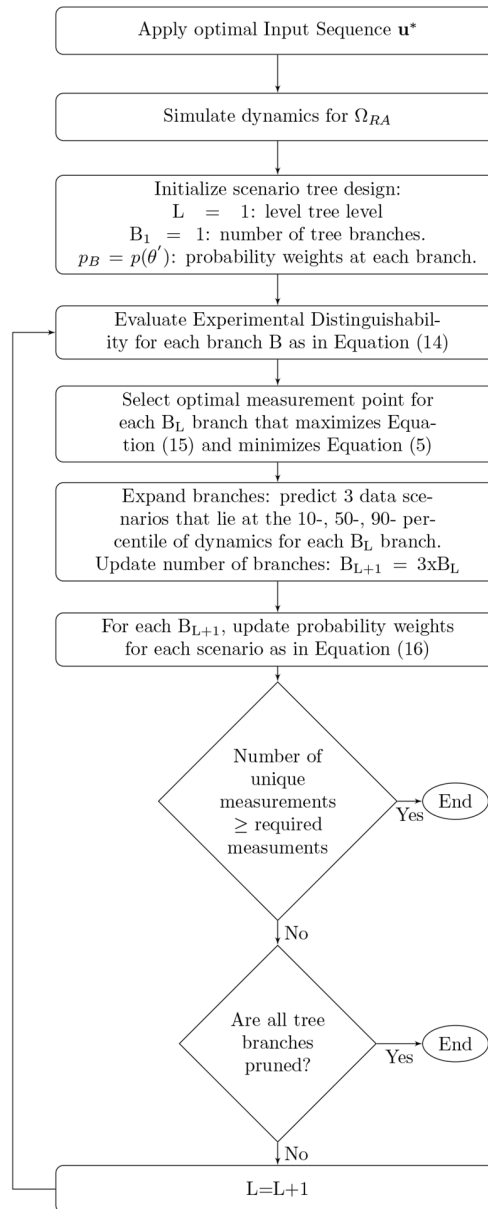
(a) Identify Representative Parameters



(b) Determine Optimal Input Vector



(c) Specify Multiple Measurement Pairs



**Fig 1. Flow charts describing the three steps for the MBDOE method.** These describe (a) how acceptable and representative parameters are identified, (b) the process for determining the optimal input vector, and (c) the selection of multiple measurement pairs associated with the optimal input vector.

doi:10.1371/journal.pcbi.1004488.g001



simulations to fit the available data as:

$$\Omega_A = \left\{ \theta' \in \Omega \mid \log_{10} \left( 1 + \sum_{t \in \Omega} \left( \frac{\tilde{y}_i(\mathbf{u}, \theta', t) - \hat{y}_i(\mathbf{u}, t)}{\sigma_i(t)} \right)^2 \right) \leq T_A \right\} \quad (7)$$

where  $\hat{y}_i(\mathbf{u}, t)$  is the mean of the experimental data for the  $i^{th}$  model output at the time point,  $t$ , collected with the applied input  $\mathbf{u}$ ,  $\tilde{y}_i(\mathbf{u}, \theta', t)$  is the corresponding model dynamics simulated with a parameter set,  $\theta'$ , and  $\sigma_i(t)$  is the standard deviation of the data. In this work, we approximate  $\tilde{y}_i(\mathbf{u}, \theta', t)$  using a sparse grid interpolation tool over the uncertain space  $\Omega$  for computational efficiency.  $T_A$  is the threshold for acceptability. This weighted least squares function has been used before to define acceptable parameters in [14, 15]. The parameter space,  $\Omega$ , is sampled using Latin Hypercube Sampling (LHS). If the initial screen produces fewer acceptable parameters than  $N_A$ , the desired number of acceptable parameter vectors that will support the experiment design, focused grids are created [15] to improve the resolution of the grid interpolant in the acceptable regions of the uncertain parameter space. LHS is also used to sample the focused grids to find more acceptable parameters until the  $cardinality(\Omega_A) \geq N_A$ . In this work, we define the *dynamical uncertainty region* to be the region spanned by the most extreme trajectories of the acceptable parameters.

**Selection of Representative Parameters.** We select representative parameters from the acceptable parameter space,  $\Omega_{RA} \subset \Omega_A$ , that span the dynamics space, denoted by  $\mathfrak{D}_A$ , simulated by the model over the acceptable parameter space. Our selection of  $\Omega_{RA}$  effectively samples the parameters consistent with the number of experimentally distinguishable dynamics for robustness. We cluster the trajectories,  $\eta \in \mathfrak{D}_A$ , where  $\eta = \varphi(\theta')$  is the trajectory associated with  $\theta'$ , so that parameters that generate similar dynamics are grouped together. A spectral clustering method is used that employs eigenvalue analysis to estimate the number of clusters. The cluster number is determined at the largest gap of the eigenvalues of the normalized Laplacian matrix that is used in spectral clustering as in [39]. After the clusters are created, we select representative parameters from each cluster to form  $\Omega_{RA}$ . To preserve dynamical diversity, the representative parameters are sampled from the convex hull of each cluster using the MATLAB function *convhulln*. The number of representatives sampled from the convex hull of each cluster is equal to a user-specified number,  $C$ , with minimum sample size of three to ensure triangulation. Thus the representative parameters discretize the acceptable uncertain space into  $N_{RA} = C \cdot (\# \text{ of clusters})$  parameters that preserve the dynamical diversity. We use these representative parameters as an indirect means to approximate a uniform sampling from the acceptable trajectory space as described next.

**Computational Efficiency Afforded by Dynamical Representative Parameters.** For computational efficiency, our goal is to estimate the mean and variance of possible trajectories with uniform sampling in dynamical space. To be more precise, let  $\ell_{\Omega}(\theta')$  be the likelihood of  $\theta'$  based on existing experimental data. Then one way to define the probability of  $\theta'$  is:

$$p_{\Omega}(\theta') = \frac{\ell_{\Omega}(\theta')}{\sum_{\lambda \in \Omega_A} \ell_{\Omega}(\lambda)} \quad (8)$$

However, this focuses on the parameter space and may overweigh some regions of the dynamical space if the dynamics are relatively insensitive to changes in the parameter values. This is due to the fact that we obtained  $\Omega_A$  by LHS on the parameter space. Instead, we want to sample uniformly in the dynamical space. Let  $\mathfrak{D}_{RA} = \varphi(\Omega_{RA})$  be the set of trajectories simulated with the set of representative parameters. So the probability of a given trajectory,  $\eta$ , can be defined

by:

$$p_{\mathfrak{D}_{RA}}(\eta) = \frac{\ell_{\mathfrak{D}}(\eta)}{\sum_{v \in \mathfrak{D}_{RA}} \ell_{\mathfrak{D}}(v)} \tag{9}$$

This can be used to translate to a  $p_{\Omega_{RA}}(\theta')$  where  $\theta'$  generates a trajectory,  $\eta$ , that is contained in  $\mathfrak{D}_{RA}$ . (Recall,  $\eta \in \mathfrak{D}_A$ .) Thus, an alternative way to define the probability of  $\theta'$  over the experimentally distinguishable dynamical space is:

$$p_{\Omega_{RA}}(\theta') = \frac{\ell_{\Omega}(\theta')}{\sum_{\beta \in \Omega_{RA}} \ell_{\mathfrak{D}}(\beta)} \tag{10}$$

Since  $\mathfrak{D}_{RA}$  spans the dynamical set represented within  $\mathfrak{D}_A$ , its reduced membership is compensated for by these probabilities.

The representative acceptable parameter set and its associated probability weights enable us to estimate the expected values and variances for the measured and target states over the experimentally distinguishable dynamics as:

$$\sum_{\mathfrak{D}_A} (f) \approx \sum_{\eta \in \mathfrak{D}_{RA}} p_{\mathfrak{D}_{RA}}(\eta) f(\eta) = \sum_{\theta' \in \Omega_{RA}} p_{\Omega_{RA}}(\theta') f(\theta') \tag{11}$$

Table 1 specifies where this relationship in Eq (11) is used in the following equations to estimate the expected values and variances in a computationally efficient manner using the representative parameters.

### Determine Optimal Input Vector

We propose solving the input vector using a greedy search algorithm. The input is discretized as in control vector parameterization (CVP) by the potential admissible input times  $\mathcal{T} = \{\tau_1 \dots \tau_N\}$  and by potential magnitude levels. The number of possible input magnitudes is determined by the inputs bound  $[u_{min}, u_{max}]$  and resolution,  $\delta u$ :

$$N_u = \frac{u_{max} - u_{min}}{\delta u} + 1 \tag{12}$$

The algorithm for the greedy search method to select the optimal input vector is detailed in the flowchart in Fig 1(b). The input magnitude is initially set to a base input level,  $\mathbf{u}_1 = \mathbf{u}_b(t) \forall t \in [T_I, T_F]$ , which is determined by the user. At each iteration,  $k$ , of the greedy search method, optimized input levels,  $\mathbf{u}_k^*$ , for the admissible times,  $\tau_j \in \mathcal{T}_{k-1}$ , are selected from all  $N_u$  possible

**Table 1. Mapping of functions to equation numbers to efficiently estimate necessary expected values and variances.**

Equation number	Expected Value Description	f(θ)
(14)	measurement variance, $\text{Var}(y_i(\mathbf{u}, \cdot, t))$	$(y_i(\mathbf{u}, \theta', t) - \mu_i(t))^2$
(14) and (18)	expected experimental noise*, $E(\sigma_i(\mathbf{u}, t))^2$	$\zeta_b + \zeta_s   y_i(\mathbf{u}, \theta', t)   + \zeta_t   \dot{y}_i(\mathbf{u}, \theta', t)  $
(5) and (18)	mean value, $\mu$	$y_i(\mathbf{u}, \theta', t_k)$
(5)	target state variance, $\text{Var}(x_{T_i}(t))$	$(x_{T_i}(\mathbf{u}_T, \theta', t) - \mu_{T_i}(t))^2$

\*the constants  $\zeta_b, \zeta_s, \zeta_t$  are described in [12].

doi:10.1371/journal.pcbi.1004488.t001



admissible input magnitudes as described in Algorithm 1. We evaluate  $TDU^j$  for each of the admissible input times  $\tau_j \in \mathcal{T}_{k-1}$  and associated  $\mathbf{u}_j^*$  and select the input conditions,  $(\tau_j^*, \mathbf{u}_j^*)$ , that minimize the value of TDU according to Eq (6) as optimal. These optimal input conditions update the current input vector as follows:

$$\mathbf{u}_k^* = \begin{cases} \mathbf{u}_j^* & \tau_j^* \leq t \leq \tau' \\ \mathbf{u}_{k-1}^* & \text{Elsewhere} \end{cases} \quad (13)$$

where  $\tau'$  is an input admissible time that has previously been optimized to update the vector,  $\mathbf{u}_{k-1}^*$ . We define  $\mathcal{S}_{k-1}$ , a set of all previously optimized input admissible times such that  $\tau' = \min[t \in \mathcal{S}_{k-1} \mid t > \tau_j^*]$ . When the input vector is updated with optimal input conditions,  $(\tau_j^*, \mathbf{u}_j^*)$ , the input time,  $\tau_j^*$  is moved from the set  $\mathcal{T}_{k-1}$  to the set  $\mathcal{S}_{k-1}$ . This process continues iteratively, and the input vector is updated until  $\mathcal{T}$  is empty or no input magnitude change improves the value of  $TDU_{k-1}$  as shown in Fig 1(b).

**Algorithm 1** Algorithm for Selecting  $\mathbf{u}^*$

**Input:**  $\tau_j, \mathbf{u}_{min}, \mathbf{u}_{max}, \delta \mathbf{u}, \mathcal{S}_{k-1}, \mathcal{T}_{k-1}, \mathbf{u}_{k-1}^*$

**Output:**  $\mathbf{u}_j^*$  and corresponding  $TDU^j$

**for**  $p = 1 : N_u$  **do** (Note:  $N_u$  = number of possible input magnitudes)

1. Define  $\mathbf{u}_j^p = \mathbf{u}_{min} + (p - 1)\delta \mathbf{u}$
2. Simulate dynamics with  $\mathbf{u} = \mathbf{u}_p$  for  $\tau_j \leq t < \tau'$  where  $\tau' = \min[t \in \mathcal{S}_{k-1} \mid t > \tau_j]$
3. Determine an associated measurement pair:  $m_k \in n_m$  and  $t_k \in \mathbb{T}$ . See Eq (15)
4. Update uncertain parameter probabilities given data for  $(m_k, t_k)$ . See Eq (16)
5. Estimate variance of target states,  $\mathbf{x}_T$ , with target input,  $\mathbf{u}_T$ , using updated uncertain parameter probabilities
6. Calculate  $TDU^p$ . See Eq (5)

**end for**

$p^* = \operatorname{argmin}_{p \in \{1, \dots, N_u\}} TDU^p$ . See Eq (6)

$\mathbf{u}_j^* = \mathbf{u}_j^{p^*}$

$TDU^j = TDU^{p^*}$

**Determine an Associated Measurement Pair.** Given an input vector,  $\mathbf{u} \in \mathbb{R}^{n_u \times N}$ , a measurement pair is selected from the feasible measurement space defined by the  $m_k^{th}$  entry in the  $\mathbf{y}$  vector at a specific sample time point,  $t_k$ . This pair is selected as the one that maximizes a Hampel smoothed distinguishability metric. We smooth the function of the distinguishability metric,  $\xi_i(t)$ , over time using a Hampel identifier,  $\Gamma_{Hampel}(\bullet)$ , to remove outliers and smooth the grid approximation (when used). The distinguishability metric quantifies the ability of an experiment to resolve the dynamics and is defined as:

$$\xi_i(t) = \Gamma_{Hampel} \left( \frac{\operatorname{Var}(y_i(\mathbf{u}, \cdot, t))}{E(\sigma_i(\mathbf{u}, t))^2} \right) \quad (14)$$

where  $\xi_i(t)$  is the output's distinguishability metric for the for the  $i^{th}$  measurable model output at each possible measurement time point.  $E(\sigma_i(\mathbf{u}, t))^2$  reflects the expected experimental noise of a measurement. If the experimental measurement noise is not available prior to running the experiments, we estimate the noise with an error model adopted from [14] as shown in

**Table 1.** The optimal measurement pair satisfies:

$$(m_k, t_k) = \operatorname{argmax}_{i \in \mathcal{N}_m, t \in \mathbb{T}} \zeta_i(t) \tag{15}$$

where  $m_k$  is the  $i^{\text{th}}$  species of  $\mathbf{y}$  and  $t_k$  is the sample time point.

This distinguishability metric is similar to the one introduced in [13] with the exception of the Hampel smoothing and the probability weights which were added to smooth the approximations for smaller numbers of sampled uncertain parameters as described in Eqs (8) to (11). The key point is that we focus on selecting measurements to distinguish dynamics rather than parameters. Hence, probability weights,  $p(\theta)$ , are induced from a probability on dynamics in a way that will allow us to sample uniformly in dynamical space. This enables us to calculate the expected values and variances in Eq (14) over the experimentally distinguishable dynamics in a computationally efficient manner.

**Update Uncertain Parameter Probabilities.** Given the optimal measurement pairs  $(m_k, t_k)$ , we use data that corresponds to the measurement point to assign and update parameter probability weights. Each of the uncertain parameters,  $\theta \in \Omega_{RA}$ , is assigned a probability weight to reflect the confidence given the simulation fitness for various scenarios. Herein, the priors of these probability weights are initially assigned uniformly across  $\theta \in \Omega_{RA}$  and updated following Bayes theorem:

$$p(\theta' | \hat{y}_i(\mathbf{u}, t_k)) = \frac{p(\hat{y}_i(\mathbf{u}, t_k) | \theta') p(\theta')}{\tilde{p}(\hat{y}_i(\mathbf{u}, t_k))} \tag{16}$$

$$\tilde{p}(\hat{y}_i(\mathbf{u}, t_k)) \approx \sum_{\theta' \in \Omega_{RA}} p(\hat{y}_i(\mathbf{u}, t_k) | \theta') p(\theta') \tag{17}$$

The posterior parameter probability,  $p(\theta' | \hat{y}_i(\mathbf{u}, t_k))$ , under the observation is given by the prior probability,  $p(\theta')$ , and the likelihood function,  $p(\hat{y}_i(\mathbf{u}, t_k) | \theta')$ , normalized by the normalizing constant,  $\tilde{p}(\hat{y}_i(\mathbf{u}, t_k))$  which is approximated in Eq (17).  $\hat{y}_i(\mathbf{u}, t_k)$  is the data of the  $i^{\text{th}}$  measurement species at the specific time point  $t_k$  collected under the application of the input  $\mathbf{u}$ . If experimental data is available, then  $\hat{y}_i(\mathbf{u}, t_k)$  is the observed data. However, for our parallel design, we make predictions of the data for the selected measurement point since data is not available until the experiment design is over. In this case,  $\hat{y}_i(\mathbf{u}, t_k)$  is the prediction of the data for the selected species at the particular time point. With the design of multiple measurements employed in this method, data is predicted with a scenario tree design as described below and the likelihood of the predicted data is assumed to follow a Gaussian distribution:

$$p(\hat{y}_i(\mathbf{u}, t_k) | \theta') = \frac{1}{\sqrt{2\pi E(\sigma_i(\mathbf{u}, t_k))^2}} \exp\left(-\frac{(\hat{y}_i(\mathbf{u}, t_k) - y_i(\mathbf{u}, \theta', t_k))^2}{2E(\sigma_i(\mathbf{u}, t_k))^2}\right) \tag{18}$$

where  $y_i(\mathbf{u}, \theta', t_k)$  is the model simulation of the dynamics of the model at the selected measurement point  $(m_k, t_k)$  with the parameter  $\theta'$  under the input  $\mathbf{u}$ .

### Specifying Multiple Measurement Pairs

Our design can identify multiple informative measurements given an optimal selected input vector,  $\mathbf{u}^*$ , as shown in Fig 1(c). We use an adaptation of the measurement scenario tree method [15] which makes multiple predictions of the data value from a given measurement point to allow the next informative point to be selected. In this work, we introduce the use of prior/posterior probability updates to minimize the bias due to outliers. The use of probabilities also enables

the algorithm to work well with a smaller number of sampled uncertain parameters than previously, contributing to the computational efficiency. As in [15], a measurement scenario tree is initialized by a node which defines the first optimal measurement point. At each node, we determine the optimal sampling time point according to Eq (15) to maximize the distinguishability metric for each measurable species. Our MBDOE algorithm selects the measurement pair that minimizes the TDU according to Eq (5) as the optimal measurement pair for that node. Subsequent measurements (nodes) are chosen by considering three different scenarios (branches) from the node. These scenarios arise by predicting three possible outcomes for the measurement pair: previously these were defined by the minimum, mean, and maximum predicted measurement values. Herein, the cumulative distribution function is computed from the priors to span from the minimum to maximum predicted data values for that measurement pair. The three predicted possible outcomes are the simulated values at the measurement pairs that correspond to the 10<sup>th</sup>, 50<sup>th</sup>, and 90<sup>th</sup> percentiles. Using each of these predicted possible outcomes to approximate the data,  $\hat{y}_i(\mathbf{u}, t_k)$ , the posterior for each branch are updated for the  $\theta \in \Omega_{RA}$  as in Eq (16). Given the updated branch specific priors, the measurement pair specifying the next node in each branch is determined by the optimal measurement pair selected to minimize Eq (5) assuming the estimated posterior is the prior. This process is continued until the tree structure contains at least the specified number of unique measurements. The minimum number of levels in the tree is  $\log_3(K)+1$  where  $K$  is the user-specified desired number of measurements.

### Computational Efficiency Afforded by Sparse Grid

Evaluating all possible experiments over the design space,  $\mathbb{D}$ , using model simulations for a large uncertain parameter space,  $\Omega$ , is computationally prohibitive. The sparse grid tool offers an efficient way to approximate the system’s output dynamics using interpolating polynomials by sampling an uncertain space in a systematic way to create a surrogate model for the system. The grid approximation is then used to interpolate additional points on the uncertain space without the cost of directly simulating the model.

**Sparse grid Interpolation over Uncertain Parameter Space.** A sparse grid approximation of the dynamics of the measured states over the uncertain parameter space is created using the process described previously [15]. With the sparse grid interpolant, points are sampled from  $\Omega$  using LHS and their corresponding model output are approximated by:

$$\tilde{y}_i(\cdot, \theta', t) = L_{d,q}(t) \times \mathbf{y}_i(\cdot, \theta', \mathbf{T}_s^i), \tag{19}$$

where  $\tilde{y}_i(\cdot, \theta', t)$  is the estimated value at time  $t$  for the  $i^{th}$  model output and  $\mathbf{y}_i(\cdot, \theta', \mathbf{T}_s^i)$  is a vector of the  $i^{th}$  model output sampled at time points,  $\mathbf{T}_s^i$ , specified by the time-interpolating vector for the uncertain parameters,  $\theta'$ . Herein, the  $L_{d,q}$  are Lagrange polynomials where the degree  $d$  can be increased to improve the accuracy of the interpolant approximation while ensuring computational efficiency and  $q$  is used as the index to build the polynomial,  $0 < q < d$ .

**Sparse grid Interpolation over Input Space.** If the number of all possible input combinations exceed the maximum number of grid nodes predicted to create the sparse grid interpolation, the sparse grid interpolation tool is used to approximate the model dynamics over the input space. The sparse grid interpolant is sampled at the  $N_u$  possible input magnitudes and the corresponding model outputs are approximated for each  $\theta \in \Omega_{RA}$  by:

$$\tilde{y}_i(\mathbf{u}, \cdot, t) = L_{d,q}(t) \times \mathbf{y}_i(\mathbf{u}, \cdot, \mathbf{T}_s^i), \tag{20}$$

The interpolated dynamics are used to characterize the experimental distinguishability metric (Eq (14)) for each of the  $N_u$  possible input magnitudes.

## Computational Implementation Details

The algorithm was coded in MATLAB (2013a). The Sparse Grid toolbox (v5.1.1) was obtained from <http://www.ians.uni-stuttgart.de/spinterp> [40] and integrated with our model-based experiment design algorithm. We minimize the error due to inaccurate interpolants by specifying the termination conditions based on their depth or by the relative or absolute errors of 10% and 5 respectively. The max depth is set to 3 and 5 for the parameter screening and focused grids, respectively, and 5 for the input grid. We set the data-consistent acceptability threshold  $T_A = 2$ , which correspond to two standard deviations. The Hes1 and T-cell receptor models for the results were numerically integrated using the stiff ode solver ode15s with the default settings. To increase computational efficiency, the sparse grid code and the mathematical models were vectorized. The code was run on four 8-core Intel Xeon 3.4 GHz CPUs each with 16 GB of memory and running the Windows 2003 server platform with the MATLAB Parallel Computing Toolbox. Our input design MBDOE algorithm code is available upon request from the corresponding author.

An upper bound on the number of model simulations used by our MBDOE algorithm can be approximated by:

$$T_{eval} = \underbrace{SG_\theta + N_A}_{(a)} + \underbrace{\min(N_u, SG_u) \times N_{RA} \times \frac{N(N+1)}{2}}_{(b)} + \underbrace{N_{RA}}_{(c)} \quad (21)$$

where  $SG_\theta$  and  $SG_u$  are the number of model simulations (nodes) used to create sparse grid interpolants over the uncertain parameter and input space, respectively. An upper bound estimate of these values can be calculated using a Sparse Grid toolbox function, *spdim(maxDepth, Dim)* assuming the maximum possible depth and the dimension of the uncertain parameter or input space. The number of simulations required to build an accurate interpolant may be quite a bit lower than this approximation when the model is smooth over the uncertain parameter and/or input space. The terms of this upper bound approximation are mapped to the steps of Fig 1. The (a) term of Eq (21) is the total number of model simulations used to identify the representative parameters. The second term, (b), is the number of simulations that determine the optimal input vector. For this approximation, the sparse interpolant on the input space is only created if the estimated maximum number of nodes,  $SG_u$  required is less than  $N_u$ . Furthermore, we assume a worst case scenario where the input magnitude changes at all  $N$  admissible times resulting in the fraction,  $\frac{N(N+1)}{2}$ . The third term, (c), is the number of model simulations used create the scenario tree to determine the measurement pairs.

## Results

### 3-Dimensional Problem: The Hes1 Model

To illustrate the effectiveness of our MBDOE strategy to resolve dynamical uncertainty of a target system, we use a simple Hes1 oscillator toy model [41]. This example demonstrates how an optimal perturbation enhances the ability of the MBDOE to reduce the dynamical uncertainty. The Hes1 model was previously used to demonstrate a Bayesian design of experiments strategy [16]. This model describes the changes in the level of the Hes1 transcription factor that is important in somitogenesis. It is modeled with an ODE system that describes the changes of the levels of the Hes1 mRNA,  $m$ , the cytosolic protein,  $P_1$ , and the nuclear protein,  $P_2$ , as

**Table 2. Hes1 model parameters and definitions.**

Symbol	Parameter Description	Nominal value	Parameter Range
$k$	degradation rate	$0.03 \text{ min}^{-1}$	fixed
$k_1$	transport rate of protein	$0.16 \text{ min}^{-1}$	fixed
$P_0$	regulatory transcription level of protein	2.4	0.24–24
$h$	Hill type coefficient	2	0.2–20
$\nu$	rate of mRNA transcription	$2.5 \times 10^{-2} \text{ min}^{-1}$	$2.5 \times 10^{-3}$ – $2.5 \times 10^{-1}$

doi:10.1371/journal.pcbi.1004488.t002

shown:

$$\begin{aligned} \frac{dm}{dt} &= -km + \frac{1}{1 + \left(\frac{P_2}{P_0}\right)^h}, \\ \frac{dP_1}{dt} &= -kP_1 + \nu m - k_1P_1, \\ \frac{dP_2}{dt} &= -kP_2 + k_1P_1 \end{aligned} \tag{22}$$

A description of the parameters of the model and their nominal values are listed in [Table 2](#). The degradation rate,  $k$ , and protein transport rate,  $k_1$ , are assumed to be known and their values are set to the nominal value while the rest of the parameters are considered unknown. The uncertain parameter space,  $\Omega$ , is initially set to have bounds of 0.1 and 10 times previously reported nominal parameter values.

The experimental design space is  $\mathbb{D}$ , partially defined by an input  $\alpha \in \mathbb{U}$  that modifies the transport rate,  $k_1$ , as  $\alpha \times k_1$  where  $\mathbb{U} = [0.01, 2]$  with a  $\delta u = 0.05$ . The input magnitude can be changed at admissible times,  $\mathcal{T} = \{0, 2, 5, 8, 10, 50, 100, 150, 200\}$  min. In this example,  $u_b$  is set to  $\alpha = 1 \forall \tau \in \mathcal{T}$ . The allowable measurement state space includes the Hes1 mRNA,  $m$ , and the total protein concentration,  $P_1 + P_2$ . The allowable sampling time space  $\mathbb{T} = [0, 300]$  min was defined to specify measurements with a  $\delta t = 10$  min. We want the algorithm to specify a minimum of 8 distinct measurements. The experiment is designed to decrease the dynamical uncertainty on three target states,  $m$ ,  $P_1$ , and  $P_2$  under a target input  $\alpha = 1$  applied for the simulated time period.

To evaluate the MBDOE algorithm *in silico*, we simulate a plant model for which we want to reduce the dynamical uncertainty. The plant model is initially simulated with unknown model parameters set at the nominal values shown in [Table 2](#). We generate 6 initial measurements of both mRNA,  $m$ , and total protein concentration,  $P_1 + P_2$ , at times  $t = 10, 20, 60$  min with 10% additive Gaussian noise.

Initial data identify the acceptable parameter space,  $\Omega_A$ , that is consistent with the data as defined by [Eq \(7\)](#). The resulting optimal experiment design is in [Table 3](#). The optimal input sequence is specified by  $\alpha$  values shown in [Fig 2a](#) and associated predicted measurements are shown on [Fig 2b](#), superimposed on the representative dynamics for the measurable species to illustrate their connection with uncertainty. The greedy algorithm for the input selection is shown in [Table 4](#) showing each iteration of input selection with the progressive predicted reduction in TDU values. The initial uncertainty and final uncertainty in the dynamics of the target states are superimposed in [Fig 2c](#) to demonstrate the reduction of the region of uncertainty of the dynamics. The dynamical uncertainty regions associated with the mRNA,  $P_1$ ,  $P_2$

**Table 3. Optimal Experiment Design for Hes1 Example.**

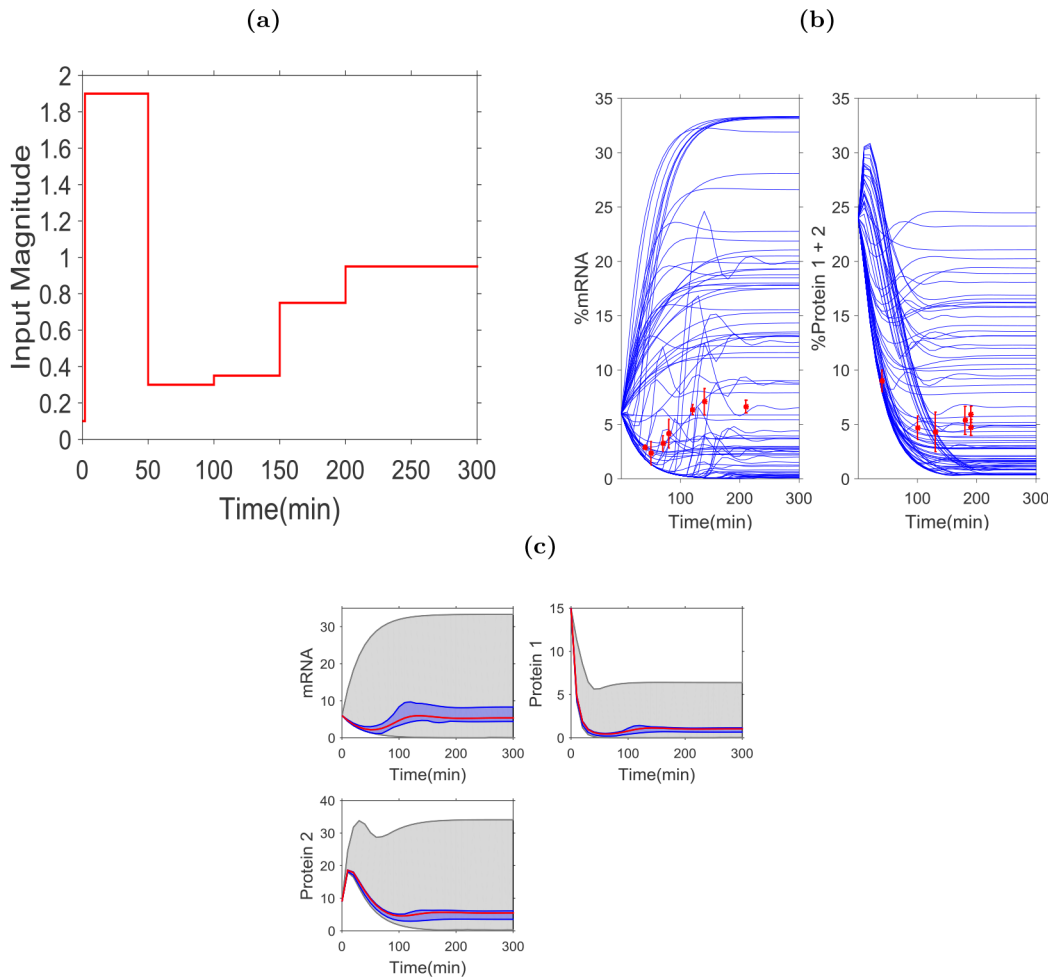
Time Step	Action under MBDOE design	Scenario Tree Level
T = 0 mins	Apply $\alpha = 0.1$	-
T = 2 mins	Apply $\alpha = 1.9$	-
	Measure P1+P2	2 <sup>nd</sup>
T = 40 mins	Measure mRNA	3 <sup>rd</sup>
	Apply $\alpha = 0.3$	-
T = 50 mins	Measure mRNA	3 <sup>rd</sup>
T = 70 mins	Measure mRNA	3 <sup>rd</sup>
T = 80 mins	Measure mRNA	3 <sup>rd</sup>
	Apply $\alpha = 0.35$	-
T = 100 mins	Measure P1+P2	3 <sup>rd</sup>
T = 120 mins	Measure mRNA	1 <sup>st</sup>
T = 130 mins	Measure P1+P2	2 <sup>nd</sup>
T = 140 mins	Measure mRNA	3 <sup>rd</sup>
T = 150 mins	Apply $\alpha = 0.75$	-
T = 180 mins	Measure P1+P2	3 <sup>rd</sup>
	Measure P1+P2	2 <sup>nd</sup>
T = 190 mins	Measure P1+P2	3 <sup>rd</sup>
T = 200 mins	Apply $\alpha = 0.95$	-
T = 210 mins	Measure mRNA	3 <sup>rd</sup>

doi:10.1371/journal.pcbi.1004488.t003

target states are reduced by 84%, 81% and 86%, respectively. In all cases, this reduced uncertainty region enclose the true dynamics of the system.

The measurement pairs are shown in the scenario tree in Fig 3a. the nodes of the tree specify the measurements that maximize uncertainty in the measurable species,  $\xi$ , while minimizing the TDU,  $\gamma$ , under the optimal input sequence,  $\mathbf{u}^*$ . In general, the values of  $\xi$  and  $\gamma$  decrease as you move down the tree and the predicted dynamics are constrained. To investigate how the measurement pairs contribute to the reduction of uncertainty in the dynamics of the target system, the total,  $\gamma$ , and individual variances are estimated assuming each predicted measurement has been taken (moving down and from left to right along the tree, neglecting repeated measurements). These results are shown in Fig 3b and 3c. These figures show that the first five unique measurements from the tree are necessary to resolve the target system dynamics.

To further explore the optimality of our MBDOE algorithm, we compare our results achieved with  $\mathbf{u}^*$  and  $M^*$  to those of measurement only MBDOE. For this comparison, the measurement only MBDOE determined the optimal measurements,  $M_T$ , assuming the applied input was the target input,  $\mathbf{u}_T$ . These results are summarized in Table 5 by the percentage reduction in the uncertain dynamical region for each target species and TDU. Our MBDOE algorithm outperforms all other experiment design combinations based on the TDU reduction. Although our MBDOE design is not the best in reducing the uncertain dynamical region for all target states, the results are comparable. The supplemental material contains the figures showing the uncertain dynamical regions for all measurement and input combinations in the Table. The supplemental material also contains figures that indicate the ability of our MBDOE algorithm to reduce the uncertain dynamical regions for different parameterizations of the Hes1 plant model.



**Fig 2. An illustration of the optimal experiment design for the Hes1 example.** (a) The input sequence selected by the design for the duration of the experiment. (b) Model simulation of the output dynamics of the measurable species  $m$  and  $P_1+P_2$  under an optimal input simulated with  $\Omega_{RA}$ . The red dots specify the selected optimal measurement points which are the mean values generated using nominal plant with additive 10% Gaussian noise. Error bars are also given by the standard deviation of the simulated data (three data points are generated for each time point). (c) The initial (grey) and final (blue) uncertainty in the target states dynamics. The designed experiment reduces the uncertainty region by 84%, 81%, and 86% for mRNA,  $P_1$ , and  $P_2$ , respectively. The red line shows the simulated dynamics of the nominal plant.

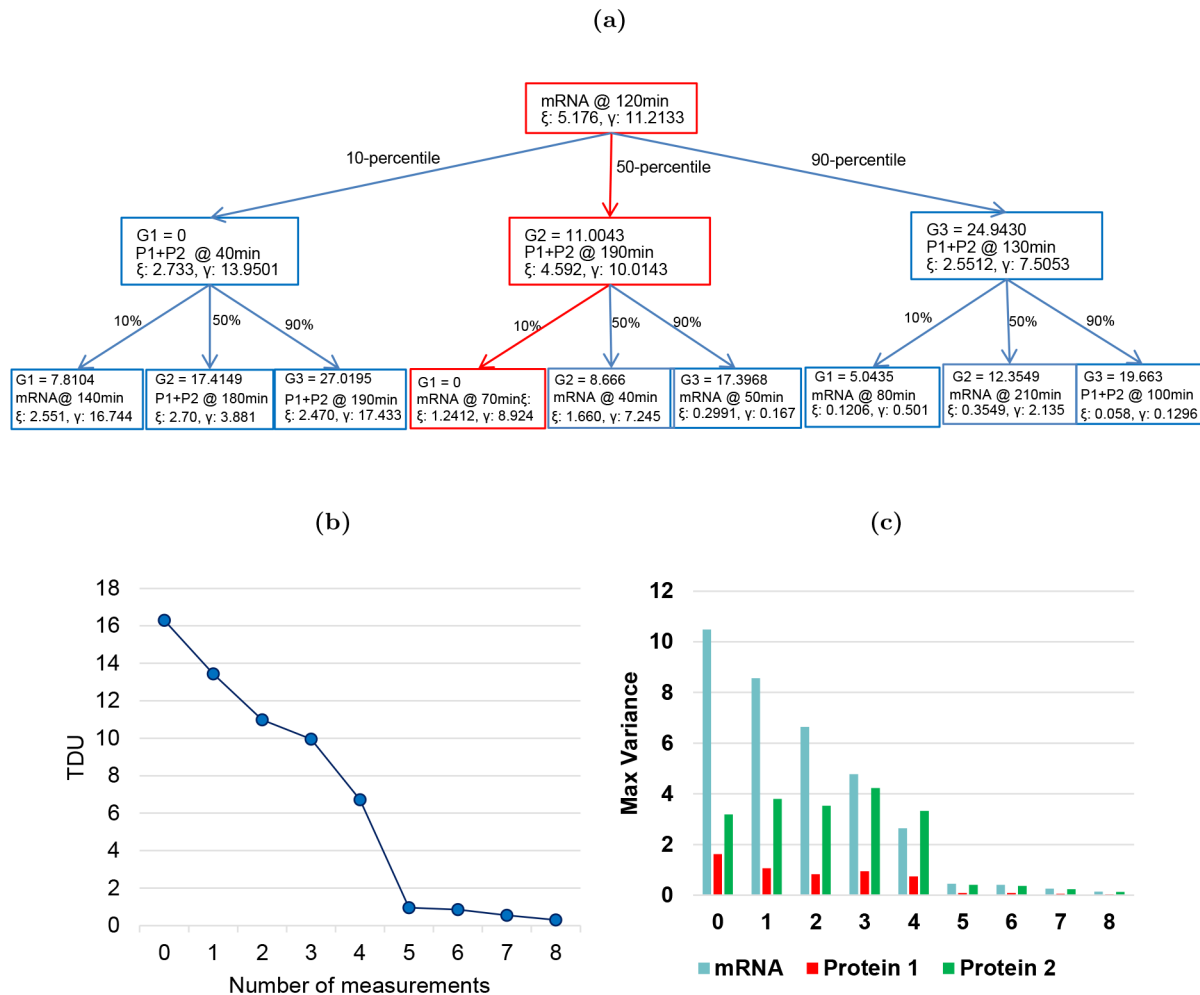
doi:10.1371/journal.pcbi.1004488.g002

**Table 4. Greedy Method for Hes1 Input Vector Design.**

Iteration	min(TDU)	$\tau^*$	$u^*$	$u$ at time $T = \{0 \ 2 \ 5 \ 8 \ 10 \ 50 \ 100 \ 150 \ 200\}$
0	–	–	–	[1 1 1 1 1 1 1 1 1]
1	11.5191	50	0.3	[1 1 1 1 1 0.3 0.3 0.3 0.3]
2	10.7718	0	0.1	[0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.3 0.3 0.3 0.3]
3	7.8597	2	1.9	[0.1 1.9 1.9 1.9 1.9 1.9 1.9 0.3 0.3 0.3 0.3]
4	7.6381	100	0.35	[0.1 1.9 1.9 1.9 1.9 1.9 1.9 0.3 0.35 0.35 0.35]
5	1.1810	150	0.75	[0.1 1.9 1.9 1.9 1.9 1.9 1.9 0.3 0.35 0.75 0.75]
6	0.5849	200	0.95	[0.1 1.9 1.9 1.9 1.9 1.9 1.9 0.3 0.35 0.75 0.95]

doi:10.1371/journal.pcbi.1004488.t004





**Fig 3. Supporting details of optimal experiment design for Hes1 example.** (a) The measurement scenario tree representation for selecting optimal measurements. Each node defines, the measurement pairs with the predicted value of the distinguishability metric,  $\xi$ , target dynamical uncertainty,  $\gamma$ , and three estimated measurement values from simulation, G1, G2, and G3. The path along the scenario tree with predictions closest to the data from the nominal plant is shown in red. (b) Reduction of TDU and (c) individual variance of target states with each additional unique measurement from the scenario tree.

doi:10.1371/journal.pcbi.1004488.g003

**Table 5. Comparison of Optimal Input MBDOE with Measurement Only MBDOE for the Hes1 Model.**

Input Applied	Measurements	% Uncertainty Reduction			
		$m$	$P_1$	$P_2$	TDU, $\gamma$
$u^*$	$M^*$	84%	81%	86%	84.6%
$u_T$	$M_T$	75%	68%	81%	77%
$u^*$	$M_T$	86%	82%	83%	84.2%
$u_T$	$M^*$	80%	79%	82%	80.8%

doi:10.1371/journal.pcbi.1004488.t005

The computational efficiency of this MBDOE method partially arises from the use of an interpolation as a surrogate of the model to reduce numerical integration of ODEs. The sparse grid terminated with a relative accuracy of 0.7% with an absolute tolerance of 0.9. The algorithm required only 137 model evaluations to build an interpolant which was sampled 10,000 times to identify 3852 acceptable parameters where  $N_A = 2000$ . The algorithm selected 48 representative parameters,  $N_{RA} = 48$ , to span the distinguishable dynamics generated by the acceptable parameters. For the input design, separate 1-D sparse grids are constructed for each admissible time of input change with 9 nodes. The total number model evaluation used for the complete experiment design is 18,716 where 2137 were for finding the representative parameters, 16,531 were used for the input selection, and 48 were used for the measurement selection. Clearly the majority of the model simulations were used to determine the optimal input signal. This is exceptionally large since we allowed the input to change at 9 time points and had fine input resolution with 40 different values. For comparison purposes, the Hes1 model experiment design in [16] required more than 30,000 model simulations to determine which species should be measured to provide the most information.

**19-Dimensional Example: The T-cell Receptor Model.** The scalability of the MBDOE strategy to high dimensional problems is demonstrated using a T-cell receptor signaling model proposed by Lipniacki et al. [42]. This model was used previously as an example model to resolve dynamical uncertainty using parallel measurements only [15]. The model has 37 ODE equations that describes the dynamics of the species involved in the activation of T-cell signaling pathway, and 19 model parameters. Herein, we assume all of the parameters are unknown.

The Lipniacki model demonstrates the effectiveness of the MBDOE algorithm to use an optimal input to reduce dynamical uncertainty in a target system with conditions that maybe difficult to manipulate experimentally exactly. For illustrative purposes, we consider the system that could be experimentally determined to have a concentration of the agonist ligand, pMHC1, significantly higher than the concentration of the TCR receptors. The target system assumes that the concentration of this ligand pMHC1 predictably decreases as it binds to the TCR receptor. Two inputs are allowed to be manipulated in the experiment design under the system with constant supersaturating concentration of pMHC1: the drugs Sanguinarine, an Erk inhibitor, and PMA, an upstream activator of the protein kinase C. The magnitude levels of these control inputs,  $\mathbb{U}$ , where  $\mathbb{U} \in [0, 1]$  with a resolution of  $\delta u = 0.3$  can be changed at potential times  $\mathcal{T} = \{5, 10, 15, 20\}$  min while the duration of an experiment run is 30 minutes. In this example, the initial  $u_b$  is set to no drug application. None of the drugs are applied to the target system.

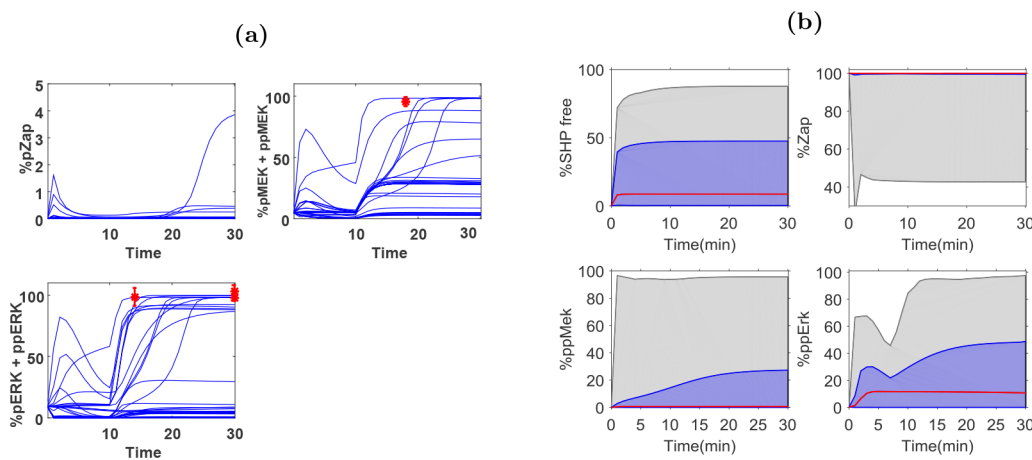
The allowable measurement state space  $\mathbb{M}$  is defined to include phosphorylated Zap, pZAP, total phosphorylated Mek, pMEK + ppMEK, and total phosphorylated Erk, pERK + ppERK. The allowable sampling time space  $\mathbb{T} = [0, 30]$  min was defined to specify measurements with a  $\delta t = 1$  min. Optimal measurements are designed under an optimal input to reduce dynamical uncertainty in target states, free SHP, ZAP, ppMEK, and ppERK under the target system with predictable decreasing levels of pMHC1. A total of four measurements are to be designed. The design is initiated by identifying acceptable parameters that fit an initial data set containing 9 points, pERK + ppERK at 0, 2, 5, 7, 8, 10, 12, 20 and 30 minutes.

The optimal experimental design that constrains the predicted dynamics of the target system is presented in Table 6. Optimal drug levels for Sanguinarine and PMA are administered at times  $t = 10$  minutes and optimal measurements are specified at times  $t = 14, 18$  and 30 minutes. These predicted measurements are also superimposed on the simulation of the representative parameters in Fig 4a to illustrate their connection with uncertainty. The optimal experimental design yields a reduction of uncertainty in target state dynamics by as much as 99% observed in Zap, with the least predicted dynamical uncertainty reduction by 46% for free

**Table 6. Optimal Experiment Design for TCR Example.**

Time Step	Action under MBDOE design	Scenario Tree Level
	Sanguinarine = 0	-
T = 10 mins	Apply PMA = 0.3	-
T = 14 mins	Measure pERK+ppERK	2 <sup>nd</sup>
T = 18 mins	Measure pMEK+ppMEK	2 <sup>nd</sup>
T = 30 mins	Measure pERK+ppERK	1 <sup>st</sup>
T = 30 mins	Measure pERK+ppERK	2 <sup>nd</sup>

doi:10.1371/journal.pcbi.1004488.t006



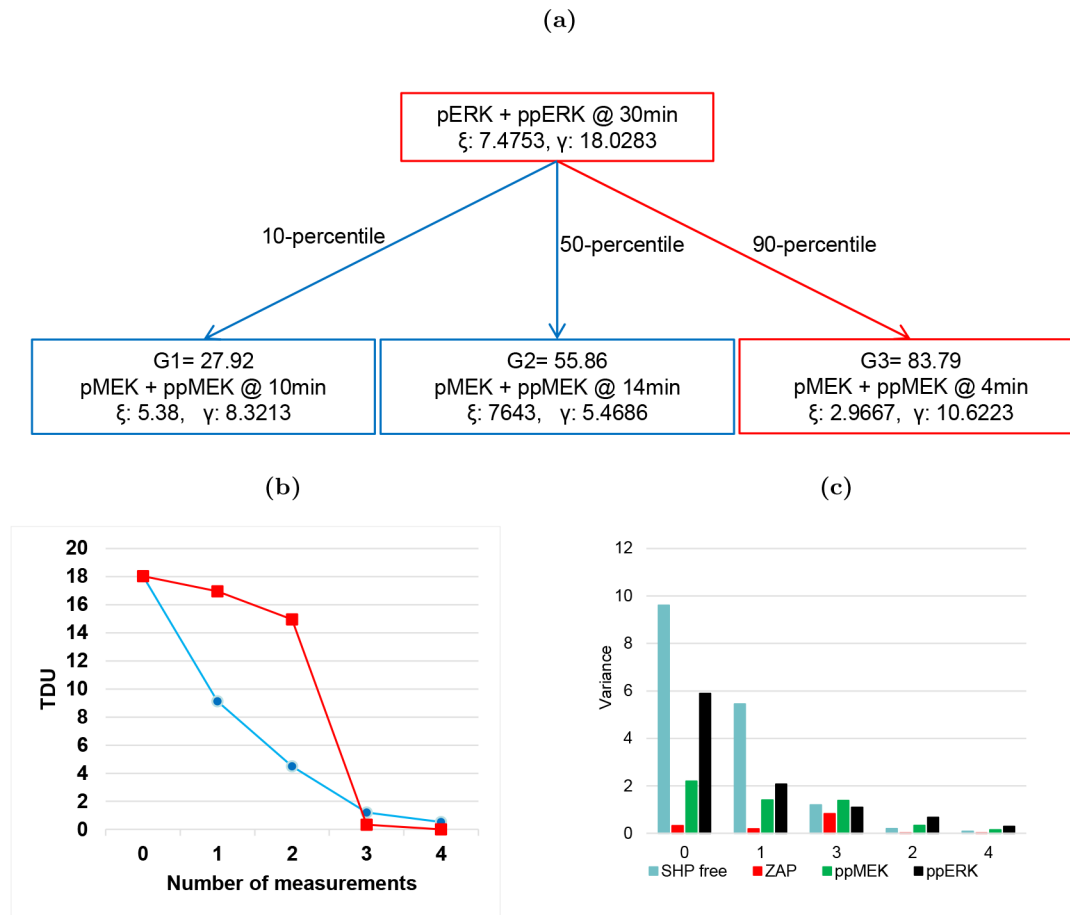
**Fig 4. An illustration of the optimal experimental design for the TCR example.** (a) Model simulation of the output dynamics of the measurable species under the optimal input provided in Table 6 simulated with  $\Omega_{RA}$ . The red dots specify the selected optimal measurement points by the mean values generated by nominal plant with additive 10% Gaussian noise. Error bars are also given by the standard deviation of the simulated data (three data points are generated for each time point). (b) The initial (grey) and final (blue) uncertainty in the target states dynamics. The designed experiment reduces the uncertainty region by 46%, 99%, 85% and 59% for free SHP, ZAP, ppMEK and ppERK, respectively. The red line shows the simulated dynamics of the nominal plant.

doi:10.1371/journal.pcbi.1004488.g004

SHP (Fig 4b). Although we would like to dramatically reduce the uncertainty for all target states, the reduction performance is highly dependent on the relationship between the target and measurable states. There are cases where the measurements may not be informative at all; our MBDOE approach will help inform the experimentalist *a priori*.

For completeness, the measurement scenario tree design is provided in Fig 5a. Each measurement from the scenario tree is investigated to determine how it contributes to the decrease in the TDU and the individual maximum variance (Fig 5b and 5c). Fig 5b also compares the reduction in TDU for our MBDOE approach to a measurement only design. For the measurement only MBDOE approach, we choose to find 4 measurements assuming the target input was applied to the TCR model. From this figure, it is clear that the optimization of the input signal is most important when a limited number of measurements are taken.

The computational efficiency of this MBDOE method on a higher dimension problem depends on the use of interpolation to explore the uncertain parameter space. From the 10,063 nodes used to create the grid, the algorithm identified 3974 acceptable parameters. The grid interpolant was sampled 100,000 times using LHS to identify an additional 403 acceptable parameters. The algorithm determined  $N_{RA} = 44$  with  $N_A = 2000$ . There was no need to



**Fig 5. Supporting details of optimal experiment design for TCR example.** (a) The measurement scenario tree representation for selecting optimal measurements. Each node defines, the measurement pairs with the predicted value of the distinguishability metric,  $\xi$ , target dynamical uncertainty,  $\gamma$ , and three estimated measurement values from simulation, G1, G2, and G3. The path along the scenario tree with predictions closest to the data from the nominal plant is shown in red. (b) Reduction of TDU (blue line) and (c) individual variance of target states with each additional unique measurement from the scenario tree. In (b) the reduction in TDU for a measurement only MBDOE scenario (red line) is provided for comparison purposes.

doi:10.1371/journal.pcbi.1004488.g005

construct an input sparse grid since  $N_u = 16$  and there were only 4 admissible times for potential input changes. The total number of model simulations for the complete experiment design is 16,991 with 12,063 for the selection of representative parameters, 4884 simulations for the input design, and 44 simulations to support the measurement selection. The majority of the model simulations are used to explore the large dimension uncertain parameter space to identify the acceptable and representative parameters. Given Eq (21) and the settings specified for the TCR model, the maximum number of model simulations possible was 19,084. The number of simulations actually performed was smaller than the worst case scenario because the input design terminated due to a lack of improvement in the TDU with additional changes to the input vector. For comparison purposes, previous work with this TCR model assuming different initial data and measurable species in [15] required 19,000 model evaluations to only define optimal measurements for a parallel design that fully resolved the dynamics to within experimental capabilities.

## Discussion

In this work, we have proposed an MBDOE algorithm that extends previous work to include the design of an optimal input together with optimal measurements to reduce dynamical uncertainty in biological systems. The global nature of the algorithm overcomes the challenges posed by traditional MBDOE designs that rely on the local FIM design. Furthermore, most MBDOE techniques that consider input design only compare a predefined set of input values and/or measurements. Our computationally efficient approach enables us to avoid this limitation and search over the input space for the optimal input values at pre-specified time points. Our strategy achieves a global design by implementing computationally efficient strategies using sparse grid interpolation, probability-weighted scenario trees, and dynamical representative parameters. An interpolant of the ODE system is created by strategically sampling the uncertain parameter space and is used to evaluate additional points by LHS without simulating the ODEs. Representative parameters are used to span the dynamics of the acceptable parameter space. Together, the representative parameters and the sparse grid interpolation are used to efficiently screen the dynamical variance on the input space and facilitate the probability-weighted scenario tree to identify associated optimal measurements that minimize the uncertainty in the target system dynamics. We have confirmed that our optimal input MBDOE algorithm will specify experiments that are more informative in their ability to reduce dynamical uncertainty over MBDOE techniques that only specify optimal measurements using 3D and 19D models. The optimization of the input signal is most important when a limited number of measurements are taken.

The computational expense of the MBDOE method is highly dependent on the dimension of the unknown parameter space and the feasible input space. For high dimension models, a large number of model simulations may be required to explore the large dimension uncertain parameter space to identify the acceptable and representative parameters. This computational burden can be reduced by the use of the sparse grid interpolation tool if the model output is smooth over the uncertain parameter space. For large input spaces, the majority of the model simulations in the MBDOE algorithm may be used to determine the optimal input signal. The number of model simulations increases with more flexibility in the input feasible space in terms of: (1) number of inputs that could be applied, (2) number of allowable changes in the inputs magnitudes (input resolutions), and (3) the number of admissible input times. In our examples, we specified a high degree of flexibility in the feasible input space for the lower dimension Hes1 model than the high dimension TCR model. As a result, the TCR model used 12,063 model simulations for the optimal input vector design while the Hes1 system required 16,531 model simulations.

Historically, MBDOE methods have designed experiments sequentially, whereby information gained from a previous experiment is used immediately to inform the next experiment design [32, 43, 44]. This approach is prevalent with the local MBDOE strategies since it improves the accuracy of the parameter estimates that support the design of the next experiment. The sequential design becomes problematic when the number of uncertain parameters is large because a single experiment does not estimate all the parameters accurately [45]. A parallel design methodology improves the estimates of the parameters and reduces experimental costs by specifying multiple measurements. The parallel design in [15] uses scenario trees to specify all optimal measurements necessary to resolve the experimentally distinguishable dynamics. A hybrid design combines the advantage of the sequential design with the cost efficiency of the parallel design by specifying multiple experiments to perform and iterating with the results informing the next round of MBDOE. Our MBDOE design uniquely supports this hybrid design strategy. It uses a measurement scenario tree to determine a user-specified

number of measurements as opposed to the previous scenario tree approach that indicated all measurements required to fully resolve the dynamics. Herein, the number of desired measurements controls the depth of the constructed scenario tree. (If a tree is made too deep, we believe that many of the specified measurements would be sub-optimal since the actual dynamics only lie within a fraction of the predicted tree.) Thus, if the desired dynamical resolution is not achieved by one application of our MBDOE algorithm, the design can be repeated while informed with the new additional experimental data. New experimental data restricts the acceptable parameter space further to generate a new more relevant optimal experimental design.

Although our MBDOE algorithm is global, we cannot claim it is globally optimal. The algorithm makes optimal decisions within constraints. To make the problem tractable, we split the optimization problem into two sequential steps while the first specifies the optimal inputs and the second finds the associated measurement pairs. It is possible, that a better solution does exist and our algorithm misses it. Another source that may contribute to a sub-optimal result would be in the accuracy of the sparse grids. If the interpolants are not sufficiently accurate surrogates for the model responses, the results may be sub-optimal. In addition, we do not presume to have a good estimate of the number of acceptable parameter values needed to support the MBDOE since it is a function of the biological system, the experimental setup, measurement noise, the mathematical model and its uncertainty. Hence, there is no exact formula for calculating the number of acceptable parameter vectors needed to support our experiment design algorithm. If the value is too small, the algorithm output would be sub-optimal and likely not repeatable. However, if repeated runs of our MBDOE algorithm produce similar results for increasing numbers of  $N_A$  than it is likely to be sufficiently large. Thus, without putting constraints on the biological system model,  $N_A$  must be sufficiently large to cover all distinguishable dynamics so that repeated application of our MBDOE algorithm derives similar optimal experiment designs.

Overall, the proposed MBDOE strategy successfully extends previous MBDOE capabilities to design an optimal input with associated measurements that minimizes the uncertainty in the target system dynamics. The interpolation grid provides a computationally efficient way to search for both parameter fits and optimal input over large uncertain spaces. The use of representative parameters, selected to span the dynamical space of the biological model, provides a means to sample the dynamical space uniformly enhancing the computational tractability of this approach. Furthermore, the probability weighted scenario tree designs modified from [15] supports input and multiple optimal measurement pair selection with fewer sampled parameters. Although the enhancement of the discrimination ability of the experiment is somewhat model dependent, for the examples we have presented herein, we have found that the ability to specify an optimal input in addition to optimal measurements has enhanced the ability to bound the expected target system dynamics.

## Supporting Information

**S1 Text. Additional results to support the efficacy of the MBDOE algorithm to reduce dynamical uncertainty in model states.**

(PDF)

## Acknowledgments

We are grateful to M Pargett, J Perley, A Sai and the rest of the Rundell research group for their insightful discussions.



## Author Contributions

Wrote the paper: TM GTB AER.

## References

1. Hasenauer J, Waldherr S, Wagner K, Allgower F (2010) Parameter identification, experimental design and model falsification for biological network models using semidefinite programming. *IET systems biology* 4: 119–130 doi: [10.1049/iet-syb.2009.0030](https://doi.org/10.1049/iet-syb.2009.0030) PMID: [20232992](https://pubmed.ncbi.nlm.nih.gov/20232992/)
2. Bandara S, Schlder 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(2): e1000558. doi: [10.1371/journal.pcbi.1000558](https://doi.org/10.1371/journal.pcbi.1000558) PMID: [19911077](https://pubmed.ncbi.nlm.nih.gov/19911077/)
3. Steiert B, Raue A, Timmer J, Kreutz C (2012) Experimental design for parameter estimation of gene regulatory networks. *PLoS One* 7(7): e40052 doi: [10.1371/journal.pone.0040052](https://doi.org/10.1371/journal.pone.0040052) PMID: [22815723](https://pubmed.ncbi.nlm.nih.gov/22815723/)
4. Bezeau M, Endrenyi L (1986) Design of experiments for the precise estimation of dose-response parameters: the Hill equation. *Journal of Theoretical Biology* 123(4): 415–430 doi: [10.1016/S0022-5193\(86\)80211-9](https://doi.org/10.1016/S0022-5193(86)80211-9) PMID: [3657186](https://pubmed.ncbi.nlm.nih.gov/3657186/)
5. Li P, Vu Q (2013) Identification of parameter correlations for parameter estimation in dynamic biological models. *BMC systems biology* 7: 91 doi: [10.1186/1752-0509-7-91](https://doi.org/10.1186/1752-0509-7-91) PMID: [24053643](https://pubmed.ncbi.nlm.nih.gov/24053643/)
6. Transtrum M, Qiu P (2012) Optimal experiment selection for parameter estimation in biological differential equation models. *BMC bioinformatics* 13: 181 doi: [10.1186/1471-2105-13-181](https://doi.org/10.1186/1471-2105-13-181) PMID: [22838836](https://pubmed.ncbi.nlm.nih.gov/22838836/)
7. Franceschini G, Macchietto S (2003) Model-based design of experiments for parameter precision: State of the art. *Chem. Eng. Sci.* 9: 11
8. Weber P, Kramer A, Dingler C, Radde N (2012) Trajectory-oriented Bayesian experiment design versus Fisher A-optimal design: an in depth comparison study. *Journal of Computational Physics* 28(18): i535–i541
9. Lages N, Cordeiro C, Sousa Silva M, Ponces Freire A, Ferreira A (2012) Optimization of time-course experiments for kinetic model discrimination. *PLoS One* 7(3): e32749 doi: [10.1371/journal.pone.0032749](https://doi.org/10.1371/journal.pone.0032749) PMID: [22403703](https://pubmed.ncbi.nlm.nih.gov/22403703/)
10. Apgar J F, Toettcher J E, Endy D, White F M, Tidor B (2008) Stimulus design for model selection and validation in cell signaling *PLoS One computational biology* 4(2): e30 doi: [10.1371/journal.pcbi.0040030](https://doi.org/10.1371/journal.pcbi.0040030) PMID: [18282085](https://pubmed.ncbi.nlm.nih.gov/18282085/)
11. Galvanin F, Macchietto S, Bezzo F, Pado V, Marzolo V (2007) Model-Based Design of Parallel Experiments. *Ind. Eng. Chem. Res* 46: 871–882 doi: [10.1021/ie0611406](https://doi.org/10.1021/ie0611406)
12. Stegmaier J, Skanda D, Lebiecz D (2013) Robust optimal design of experiments for model discrimination using an interactive software tool. *PLoS One* 8(2): e55723 doi: [10.1371/journal.pone.0055723](https://doi.org/10.1371/journal.pone.0055723) PMID: [23390549](https://pubmed.ncbi.nlm.nih.gov/23390549/)
13. Silk D, Kirk PDW, Barnes CP, Toni T, Stumpf MPH (2014). Model Selection in Systems Biology Depends on Experimental Design. Rost B, ed. *PLoS Computational Biology*. 10(6):e1003650. doi: [10.1371/journal.pcbi.1003650](https://doi.org/10.1371/journal.pcbi.1003650) PMID: [24922483](https://pubmed.ncbi.nlm.nih.gov/24922483/)
14. Donahue M M, Buzzard G T, Rundell A E (2010) Experiment design through dynamical characterisation of non-linear systems biology models utilising sparse grids. *IET systems biology* 4(4): 249–262 doi: [10.1049/iet-syb.2009.0031](https://doi.org/10.1049/iet-syb.2009.0031) PMID: [20632775](https://pubmed.ncbi.nlm.nih.gov/20632775/)
15. Bazil J N, Buzzard G T, Rundell A E (2012) A global parallel model based design of experiments method to minimize model output uncertainty. *Bulletin of mathematical biology* 74(3): 688–716 doi: [10.1007/s11538-011-9686-9](https://doi.org/10.1007/s11538-011-9686-9) PMID: [21989566](https://pubmed.ncbi.nlm.nih.gov/21989566/)
16. Liepe J, Filippi S, Komorowski M, Stumpf M P H, (2013) Maximizing the information content of experiments in systems biology. *Bulletin of mathematical biology* 9(1): e1002888
17. Marvel S W, Williams C M (2012) Set membership experimental design for biological systems. *Journal of Computational and Graphical Statistics* 6(1): 21
18. Omony J, Mach-Aigner A R, de Graaff L H, van Straten G, van Bortel A J B (2012) Evaluation of design strategies for time course experiments in genetic networks: case study of the XlnR regulon in *Aspergillus niger*. *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM*, 9(5): 1316–1325 doi: [10.1109/TCBB.2012.59](https://doi.org/10.1109/TCBB.2012.59) PMID: [22529332](https://pubmed.ncbi.nlm.nih.gov/22529332/)
19. Kutalik Z, Cho K, Wolkenhauer O, (2004) Optimal sampling time selection for parameter estimation in dynamic pathway modeling. *Bio Systems* 9(1-3): 43–55 doi: [10.1016/j.biosystems.2004.03.007](https://doi.org/10.1016/j.biosystems.2004.03.007)
20. Maheshwari V, Rangaiah G P, Samavedham L (2012) A Novel Multi-Objective Optimization based Experimental Design and its Application for Physiological Model of Type 1 Diabetes. *Symposium on Advanced Control of Chemical Processes*.



21. Kreutz C, Timmer J (2009) Experimental Design in Systems Biology. *FEBS Journal* 276(4):923–942 doi: [10.1111/j.1742-4658.2008.06843.x](https://doi.org/10.1111/j.1742-4658.2008.06843.x) PMID: [19215298](https://pubmed.ncbi.nlm.nih.gov/19215298/)
22. Schenkendorf R, Kremling A, Mangol M (2008) Optimal Experimental Design with the sigma point method. *IET Systems Biology* 3: 10–23 doi: [10.1049/iet-syb:20080094](https://doi.org/10.1049/iet-syb:20080094)
23. Flassig R. J., and Sundmacher K. (2012). Optimal Design of Stimulus Experiments for Robust Discrimination of Biochemical Reaction Networks. *Bioinformatics* 28(23): 30893096. doi: [10.1093/bioinformatics/bts585](https://doi.org/10.1093/bioinformatics/bts585)
24. Banga J R, Versyck K, Van Impe J F (2002) Computation of Optimal Identification Experiments for Non-linear Dynamic Process Models: a Stochastic Global Optimization Approach. *Industrial & Engineering Chemistry Research* 41(10): 2425–2430 doi: [10.1021/ie010183d](https://doi.org/10.1021/ie010183d)
25. Steiert B, Raue A, Timmer J, Kreutz C (2012) Experimental design for parameter estimation of gene regulatory networks. *PloS one* 7(7): e40052 doi: [10.1371/journal.pone.0040052](https://doi.org/10.1371/journal.pone.0040052) PMID: [22815723](https://pubmed.ncbi.nlm.nih.gov/22815723/)
26. Su C, Zhang H (2011) Optimal Design of Degradation Experiment Based on Monte Carlo Simulation. *Advanced Materials Research* 211–212: 280–284 doi: [10.4028/www.scientific.net/AMR.211-212.280](https://doi.org/10.4028/www.scientific.net/AMR.211-212.280)
27. Maiwald T, Kreutz C, Pfeifer A C, Bohl S, Klingmüller U, et al. (2007) Dynamic pathway modeling: feasibility analysis and optimal experimental design. *Annals of the New York Academy of Sciences* 1115:212–220 doi: [10.1196/annals.1407.007](https://doi.org/10.1196/annals.1407.007) PMID: [18033750](https://pubmed.ncbi.nlm.nih.gov/18033750/)
28. Solonen A, Haario H, Laine M (2012) Simulation-Based Optimal Design Using a Response Variance Criterion., *Journal of Computational and Graphical Statistics* 21(1):234–252 doi: [10.1198/jcgs.2011.10070](https://doi.org/10.1198/jcgs.2011.10070)
29. Huan X, Marzouk Y (2013) Simulation-based optimal Bayesian experimental design for nonlinear systems. *Journal of Computational Physics* 232(1): 288–317 doi: [10.1016/j.jcp.2012.08.013](https://doi.org/10.1016/j.jcp.2012.08.013)
30. Ryan K J (2003) Estimating Expected Information Gains for Experimental Designs With Application to the Random Fatigue-Limit Model. *Journal of Computational and Graphical Statistics* 12: 585–603 doi: [10.1198/1061860032012](https://doi.org/10.1198/1061860032012)
31. Weber P, Kramer A, Dingler C, Radde N (2012) Trajectory-oriented Bayesian experiment design versus Fisher A-optimal design: an in depth comparison study. *Journal of Computational Physics* 28(18): i535–i541
32. Gadkar K G, Gunawan R, Doyle F J (2005) Iterative approach to model identification of biological networks. *BMC bioinformatics* 6: 155 doi: [10.1186/1471-2105-6-155](https://doi.org/10.1186/1471-2105-6-155) PMID: [15967022](https://pubmed.ncbi.nlm.nih.gov/15967022/)
33. Yue H, Brown M, He F, Jia J, Kell D B (2008) Sensitivity analysis and robust experimental design of a signal transduction pathway system. *Int. J. Chem. Kinet.* 40: 730741 doi: [10.1002/kin.20369](https://doi.org/10.1002/kin.20369)
34. Mehra R K (1974) Optimal input signals for parameter estimation in dynamic systems—Survey and new results. *IEEE Transactions on Automatic Control* 19(6): 753–768 doi: [10.1109/TAC.1974.1100701](https://doi.org/10.1109/TAC.1974.1100701)
35. Dong W, Tang X, Yu Y, Nilsen R, Kim R, et al. (2005) Systems biology of the clock in *Neurospora crassa*. *PloS one* 3(8): e3105 doi: [10.1371/journal.pone.0003105](https://doi.org/10.1371/journal.pone.0003105)
36. Mdluli T, Pargett M, Buzzard G T, Rundell A, E. (2008) Specifying Informative Experiment Stimulation Conditions for Resolving Dynamical Uncertainty in Biological Systems. *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'14)*: 298–301
37. Busetto AG, Hauser A, Krummenacher G, et al (2013). Near-optimal experimental design for model selection in systems biology. *Bioinformatics*. 29(20):2625–2632. doi: [10.1093/bioinformatics/btt436](https://doi.org/10.1093/bioinformatics/btt436) PMID: [23900189](https://pubmed.ncbi.nlm.nih.gov/23900189/)
38. Betts J T and Huffman W P (198) Mesh refinement in direct transcription methods for optimal control. *Optimal Control Application Methods* 19(1): 1–21
39. Li Z, Liu J, Chen S, and Tang X (2007). Noise robust spectral clustering. In *IEEE 11th International Conference on Computer Vision*.
40. Ferenczi I (2005) *Global Optimization using Sparse Grids*. Technische Universität München 140
41. Zeiser S, Müller J, Liebscher V (2007) Modeling the Hes1 Oscillator. *Journal of Computational Biology* 14(7): 984–1000 doi: [10.1089/cmb.2007.0029](https://doi.org/10.1089/cmb.2007.0029) PMID: [17803375](https://pubmed.ncbi.nlm.nih.gov/17803375/)
42. Lipniacki T, Hat B, Faeder JR, Hlavacek WS (2008) Stochastic effects and bistability in T cell receptor signaling. *Journal of Theoretical Biology* 254: 110122 doi: [10.1016/j.jtbi.2008.05.001](https://doi.org/10.1016/j.jtbi.2008.05.001)
43. Van Derlinden E, Bernaerts K, Van Impe J F (2010) Simultaneous versus sequential optimal experiment design for the identification of multi-parameter microbial growth kinetics as a function of temperature. *Journal of theoretical biology* 264(2): 347–55 doi: [10.1016/j.jtbi.2010.01.003](https://doi.org/10.1016/j.jtbi.2010.01.003) PMID: [20064532](https://pubmed.ncbi.nlm.nih.gov/20064532/)
44. Homologous E, Rovati I, Enrico G (1988) DESIGN: Computerized Optimization of Experimental Design for Estimating  $K_d$  and  $B_{max}$  in Ligand Binding Experiments. *Journal of theoretical biology* 249: 636–649
45. Apgar J F, Witmer D K, White F M, and Tidor B (2012) Sloppy Models, Parameter Uncertainty, and the Role of Experimental Design. *Molecular BioSystems* 6(10): 1890–1900 doi: [10.1039/b918098b](https://doi.org/10.1039/b918098b)