Peer-reviewed version available at *Processes* **2017**, *5*, 63; doi:10.3390/pr5040063

Article

Multi-Objective Optimization of Experiments using Curvature and Fisher Information Matrix

Erica Manesso [1,2,†], Srinath Sridharan [3] and Rudiyanto Gunawan [1,2,*]

1 Institute for Chemical and Bioengineering, ETH Zurich, Zurich, Switzerland
2 Swiss Institute of Bioinformatics, Lausanne, Switzerland
3 Saw Swee Hock School of Public Health, National University of Singapore, Singapore
* Correspondence: rudi.gunawan@chem.ethz.ch; Tel.: +41-44-633-2134
† Current address: Bayer AG, Germany

Abstract: The bottleneck in creating dynamic models of biological networks and processes often lies in estimating unknown kinetic model parameters from experimental data. In this regard, experimental conditions have a strong influence on parameter identifiability and should therefore be optimized to give the maximum information for parameter estimation. Existing model-based design of experiment (MBDOE) methods commonly rely on the Fisher Information Matrix (FIM) for defining a metric of data informativeness. When the model behavior is highly nonlinear, FIM-based criteria may lead to suboptimal designs since the FIM only accounts for the linear variation of the model outputs with respect to the parameters. In this work, we developed a multi-objective optimization (MOO) MBDOE, where model nonlinearity was taken into consideration through the use of curvature. The proposed MOO MBDOE involved maximizing data informativeness using a FIM-based metric and at the same time minimizing the model curvature. We demonstrated the advantages of the MOO MBDOE over existing FIM-based and other curvature-based MBDOEs in an application to the kinetic modeling of fed-batch fermentation of Baker's yeast.

Keywords: design of experiments; multi-objective optimization; Fisher information matrix; curvature; biological processes; mathematical modeling

1. Introduction

Dynamic models of biological networks and processes are often created to gain a better understanding of the system behavior. The creation of dynamic biological models requires the values of kinetic parameters, many of which are system specific and typically not known *a priori*. These parameters are commonly estimated by calibrating model simulations to the available experimental data. Such parameter fitting is known to be challenging, where there often exist multiple parameter combinations that fit the available data equally well, i.e. the model parameters are not identifiable [1–5]. While there exist a number of reasons for such lack of parameter identifiability, experimental conditions have a strong influence on this issue and thus should be carefully designed. In addition, biological experiments and data collection are often costly and time-consuming, further motivating the need for well-planned experiments that would give the maximum information when given finite resources.

Model-based design of experiment offers a means for integrating kinetic modeling with experimental efforts, as illustrated by the iterative procedure in Fig. 1. The role of the model here is to capture the knowledge and information about the system up to a given iteration. By using MBDOE, one could harness this knowledge to guide experiments in the next iteration. MBDOE techniques have been used extensively for chemical process modeling, and a review of the state of the art can be found for example in the article of Franceschini and Macchietto [6]. For the purpose of parameter estimation, experiments are generally designed to generate the richest information. In this regard, the FIM whose inverse provides an estimate of the lower bound of parameter variance-covariance [7], has been commonly used to quantify data informativeness [8–14]. In the past decade, FIM-based MBDOE methods have newfound applications in emerging areas such as systems biology [11,15–17].

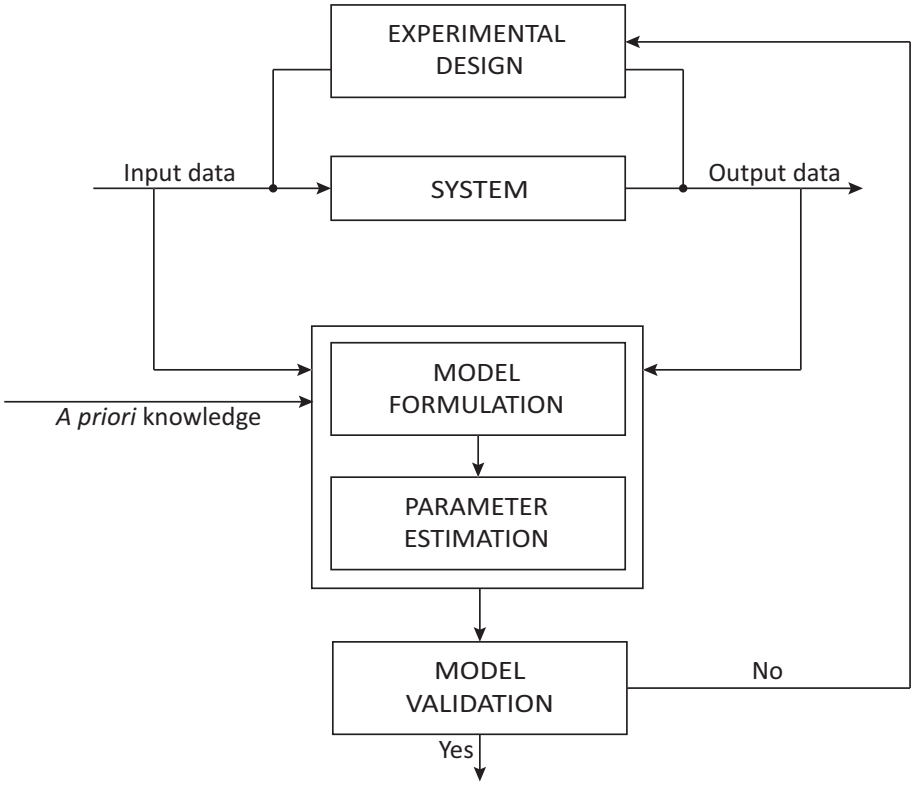

Figure 1. Iterative model identification cycle. The model building process involves the following key steps: experimental design, model structure formulation, parameter estimation, and model validation.

The FIM relies on a linear approximation of the model behavior as a function of the parameters. More precisely, the FIM is computed as a function of the first-order parametric sensitivity coefficients (Jacobian matrix) of model outputs. For systems with high degree of nonlinearity, the optimal experimental design using the FIM may perform poorly [18]. For this reason, Bates and Watts proposed a MBDOE based on minimizing model curvature by using the second-order parametric sensitivities (Hessian matrix) [19]. Hamilton and Watts further introduced a design criterion, called Q-optimality, based on a quadratic approximation of the volume of the parameter confidence region [20]. More recently, Benabbas *et al.* proposed two curvature-based MBDOEs [22]. In one design, the authors used a minimization of the root mean square (RMS) of the Hessian matrix, while in another design, they employed a constrained optimization guaranteeing the RMS to be lower than a given level. While the second strategy using a curvature threshold was demonstrated to give more informative experiments, how to set the appropriate RMS threshold value in a particular application was not described.

Recently, Maheshwari *et al.* described a MOO formulation for optimizing design of experiment using a combination of FIM-based metric and parameter correlation [17]. Since parameter correlations could not account for model nonlinearity, the strategy has the same drawback as FIM-based methods when applied to nonlinear models. In this work, we proposed a MOO MBDOE method using a combination of FIM criterion and model curvature. We demonstrated the advantages of the proposed MOO MBDOE over FIM-based and other curvature-based methods in an application to the kinetic modeling of fed-batch fermentation of Baker's yeast [21,22].

2. Model-based optimal design of experiments

We assume that the experimental data $\mathbf{y} \in \mathbb{R}^n$ are contaminated by additive random noise, as follows:

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\epsilon} \tag{1}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\epsilon}$ denote the mean of the measurement data and the random noise, respectively. When the total number of data points n is larger than the number of parameters p, $\boldsymbol{\mu}$ spans a p-dimensional space $\Omega \subset \mathbb{R}^n$ where:

$$\Omega = \{\boldsymbol{\mu} : \boldsymbol{\mu} = \mathbf{F}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\} \tag{2}$$

Here, $\mathbf{x} \in \mathbb{R}^n$ denotes the state vector, $\boldsymbol{\theta} \in \mathbb{R}^p$ denotes the parameter vector, $\mathbf{u} \in \mathbb{R}^m$ denotes the input and $\mathbf{F}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta})$ denotes the vector of non-linear model equations. The subspace Ω is also called the expectation surface or the solution *locus*. For a dynamic system, the state $\mathbf{x}$ is often described by a set of ordinary differential equations (ODEs):

$$\frac{d\mathbf{x}(\boldsymbol{\theta}, t)}{dt} = \mathbf{g}(\mathbf{x}(\boldsymbol{\theta}, t), \mathbf{u}, \boldsymbol{\theta}),\ \mathbf{x}(\boldsymbol{\theta}, 0) = \mathbf{x}_0 \tag{3}$$

The estimation of model parameters $\boldsymbol{\theta}$ from a given set of data $\mathbf{y}$ is typically formulated as a minimization of the weighted sum squares of the difference between the model prediction $\mathbf{F}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta})$ and the measurement data $\mathbf{y}$. For example, the maximum likelihood estimator (MLE) of the model parameters for normally distributed data with known variance $\mathbf{V}$, is given by the minimum of the following objective function:

$$\Phi(\boldsymbol{\theta}) = [\mathbf{y} - \mathbf{F}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta})]^T \mathbf{V}^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta})] \tag{4}$$

When the model is a linear function of the parameters: $\mathbf{F}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\theta}$, $\mathbf{X} \in \mathbb{R}^{n \times p}$, then the parameter estimates are given by $\hat{\boldsymbol{\theta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$. In this case, the MLE is the minimum variance unbiased estimator of θ, where the covariance matrix of the parameter estimates is given by $\mathbf{V}_{\boldsymbol{\theta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$. When the model is nonlinear (with respect to the parameters), the parameter estimates $\hat{\boldsymbol{\theta}} = arg\min \Phi(\boldsymbol{\theta})$ does not necessarily correspond to the minimum variance estimator. According to the Cramér-Rao inequality [7], the inverse of the Fisher Information Matrix provides a lower bound for the covariance of the parameter estimates $\hat{\boldsymbol{\theta}}$, that is:

$$\mathbf{V}_{\boldsymbol{\theta}} \geq \mathrm{FIM}^{-1} = (\hat{\mathbf{F}}^T\mathbf{V}^{-1}\hat{\mathbf{F}})^{-1} \tag{5}$$

where $\hat{\mathbf{F}} = \dot{\mathbf{F}}(\hat{\boldsymbol{\theta}}, \mathbf{x}) = \frac{\partial \mathbf{F}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ is the first order sensitivity matrix of $\mathbf{F}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta})$ with respect to the parameters θ.

Based on the Cramér-Rao inequality, the FIM has been commonly used as a criterion of data informativeness in MBDOE. Many methods for MBDOE such as those listed in Table 1, are based on finding experimental conditions that optimize a FIM-based information metric. As shown in Eq. (5), the FIM relies on a linearization of the model behavior with respect to the parameters. Essentially, the linearization replaces the expectation surface Ω by its tangent plane at $\hat{\boldsymbol{\theta}}$. The performance of the experimental design using a FIM-based criterion would therefore depend on whether: (1) the model outputs vary proportionaly with the parameter values (planar assumption); and (2) this proportionality is constant (uniform coordinate assumption) [23]. When the model is highly nonlinear with respect to the parameters, FIM-based MBDOE may produce suboptimal designs [24,25]. A recent MOO MBDOE using a combination of FIM criterion and parameter correlation, has been shown to provide improvement over FIM-based MBDOE methods[17]. But, this method also relies on the first-order parametric sensitivity matrix, and thus could not account for model nonlinearity.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2017 doi:10.20944/preprints201709.0063.v1

Peer-reviewed version available at Processes 2017, 5, 63; doi:10.3390/pr5040063

Table 1. Model-based designs of experiments using the Fisher Information Matrix (FIM).

FIM-based MBDOE	Criterion
D-optimal	max $\prod_i \lambda_i$
A-optimal	max $\sum_i \lambda_i$
E-optimal	max $min(\lambda_i)$
modified E-optimal	max $\frac{min(\lambda_i)}{max(\lambda_i)}$

Curvature-based design of experiment methods such as the Q-optimality have been introduced to account for model nonlinearity by employing a second order approximation of the model output. Here, the curvature of the expectation surface Ω is captured using the second order sensitivities of $\mathbf{F}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta})$ based on the Taylor series expansion:

$$\mathbf{F}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}) = \mathbf{F}(\mathbf{x}, \mathbf{u}, \hat{\boldsymbol{\theta}}) + \hat{\mathbf{F}}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \hat{\hat{\mathbf{F}}}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \mathrm{O}((\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^3) \tag{6}$$

where $\hat{\hat{\mathbf{F}}}_{ijk} = \frac{\partial^2 \mathrm{F}_i(\mathbf{x},\mathbf{u},\boldsymbol{\theta})}{\partial\theta_j \partial\theta_k}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ is the $n \times p \times p$ Hessian matrix. As mentioned in Introduction, several curvature-based MBDOE methods are available, for example by minimizing curvature or using curvature threshold[22]. In this work, we employed a MOO approach based on curvatures for designing optimal experiments. The basic premise of our MBDOE is to select experimental conditions which maximize the informativeness of data and ensure that the model behaves relatively linear with respect to the parameters. More specifically, our MBDOE uses two objective functions, where the first involves maximization of a FIM-based information metric, and the second involves the minimization of relative curvature measures [19]. The second objective function ensures that the FIM can provide a reliable measure of data informativeness.

2.1. Multi-objective design of experiments based on curvatures

In this section, we derive the relative curvature measures by following the work of Bates and Watts[19]. Let us consider an arbitrary straight line in the parameter space passing through $\hat{\boldsymbol{\theta}}$:

$$\boldsymbol{\theta}(b) = \hat{\boldsymbol{\theta}} + b\mathbf{h} \tag{7}$$

where $\mathbf{h} = [h_1, h_2, \ldots, h_p]$ is a non-zero vector. As the scalar parameter b is varied, a curve is traced through the expectation surface, also referred to as the lifted line, according to:

$$\boldsymbol{\mu}_\mathbf{h}(b) = \boldsymbol{\mu}(\hat{\boldsymbol{\theta}} + b\mathbf{h}) \tag{8}$$

The tangent line of this curve at b=0 is given by:

$$\begin{aligned} \boldsymbol{\mu}_\mathbf{h} &= \left[\frac{d\boldsymbol{\mu}_\mathbf{h}(b)}{db}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, b=0} \\ &= \left[\sum_{r=1}^{p} \frac{\partial \mathbf{F}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta})}{\partial\theta_r} \frac{\partial\theta_r(b)}{\partial b}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, b=0} \\ &= \hat{\mathbf{F}}\mathbf{h} \end{aligned} \tag{9}$$

The set of all such tangent lines, i.e. the column space of $\hat{\mathbf{F}}$, describes the tangent (hyper)plane at $\boldsymbol{\mu}(\hat{\boldsymbol{\theta}})$.

Meanwhile, the curvature measures come from a quadratic approximation of $\boldsymbol{\mu}$. In this case, the acceleration of $\boldsymbol{\mu}(b)$ at $b = 0$ can be written as follows:

$$\ddot{\boldsymbol{\mu}}_\mathbf{h} = \mathbf{h}^T \hat{\hat{\mathbf{F}}}\mathbf{h} = \sum_{i=1}^{p}\sum_{j=1}^{p} \frac{\partial^2 \mathbf{F}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta})}{\partial\theta_i \partial\theta_j} h_i h_j \tag{10}$$

The acceleration vector $\ddot{\boldsymbol{\mu}}_{\mathbf{h}}$ can be subsequently decomposed into two components:

$$\ddot{\boldsymbol{\mu}}_{\mathbf{h}} = \ddot{\boldsymbol{\mu}}^t_{\mathbf{h}} + \ddot{\boldsymbol{\mu}}^n_{\mathbf{h}} \tag{11}$$

where at $\boldsymbol{\mu}(\hat{\boldsymbol{\theta}})$, $\ddot{\boldsymbol{\mu}}^t_{\mathbf{h}}$ is tangential to the tangent plane and $\ddot{\boldsymbol{\mu}}^n_{\mathbf{h}}$ is normal to the tangent plane. The tangential acceleration $\ddot{\boldsymbol{\mu}}^t_{\mathbf{h}}$ is also called the parameter-effect curvature[19], which provides a measure of nonlinearity along the parameter vector $\mathbf{h}$. The degree of the parameter-effect curvature can change upon reparameterization of the model. Meanwhile, the normal acceleration $\ddot{\boldsymbol{\mu}}^n_{\mathbf{h}}$ does not vary with model parameterization, and hence is called the intrinsic curvature. Finally, the relative curvature measures in the direction of $\mathbf{h}$ are given by [19,23]:

$$K^t_{\mathbf{h}} = \frac{\|\ddot{\boldsymbol{\mu}}^t_{\mathbf{h}}\|}{\|\dot{\boldsymbol{\mu}}^t_{\mathbf{h}}\|^2} \tag{12}$$

$$K^n_{\mathbf{h}} = \frac{\|\ddot{\boldsymbol{\mu}}^n_{\mathbf{h}}\|}{\|\dot{\boldsymbol{\mu}}^n_{\mathbf{h}}\|^2} \tag{13}$$

Below, we describe the decomposition of the Hessian into the tangential and the normal component. Consider the QR factorization of the Jacobian $\hat{\mathbf{F}}$, that is $\hat{\mathbf{F}} = \mathbf{QR} = \mathbf{Q}\begin{bmatrix}\tilde{\mathbf{R}}\\ \mathbf{0}\end{bmatrix}$. By rotating the parameter axes $(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$ into $\boldsymbol{\varphi} = \tilde{\mathbf{R}}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$, a new Jacobian matrix $\dot{\mathbf{U}} = \frac{d\mathbf{F}(\mathbf{x},\mathbf{u},\boldsymbol{\varphi})}{d\boldsymbol{\varphi}}|_{\boldsymbol{\varphi}=0}$ can be computed as $\dot{\mathbf{U}} = \hat{\mathbf{F}}\tilde{\mathbf{R}}^{-1}$, which comprises the first p column vectors of $\mathbf{Q}$ (i.e. $\mathbf{Q} = \begin{bmatrix}\dot{\mathbf{U}} & \mathbf{N}\end{bmatrix}$). The remaining column vectors of $\mathbf{Q}$ (i.e. $\mathbf{N}$) are orthonormal to the tangent surface at $\boldsymbol{\varphi}$=0. In the same manner, the Hessian matrix in the rotated axes can be written as $\ddot{\mathbf{U}} = \mathbf{L}^T\hat{\mathbf{F}}\mathbf{L}$, where $\mathbf{L} = \tilde{\mathbf{R}}^{-1}$ and $\ddot{U}_{ijk} = \frac{\partial^2 F_i(\mathbf{x},\mathbf{u},\boldsymbol{\varphi})}{\partial\varphi_j\varphi_k}|_{\boldsymbol{\varphi}=0}$. The decomposition of the Hessian into the tangential and normal components is given by the following equation[19]:

$$\ddot{\mathbf{A}} = \mathbf{Q}^T\ddot{\mathbf{U}} = \begin{bmatrix}\dot{\mathbf{U}} & \mathbf{N}\end{bmatrix}^T\ddot{\mathbf{U}} = \begin{bmatrix}\ddot{\mathbf{A}}^t & \ddot{\mathbf{A}}^n\end{bmatrix} \tag{14}$$

The matrices $\ddot{\mathbf{A}}^t$ and $\ddot{\mathbf{A}}^n$ respectively correspond to the parameter-effect and intrinsic curvature components of the Hessian.

To normalize the relative curvatures in Eqs. 12 and 13, Bates and Watts[19] used the scaling factor ρ, where $\rho = s\sqrt{p}$ and $s^2 = \frac{(\mathbf{y}-\hat{\boldsymbol{\mu}})^T(\mathbf{y}-\hat{\boldsymbol{\mu}})}{n-p}$. Following the same procedure, we define the normalized relative curvatures as follows:

$$\gamma^t_{\mathbf{h}} = \rho K^t_{\mathbf{h}} \tag{15}$$

$$\gamma^n_{\mathbf{h}} = \rho K^n_{\mathbf{h}} \tag{16}$$

In addition, recasting $\mathbf{h}$ in the rotated axes as $\mathbf{h} = \mathbf{Ld}$, the tangent line $\dot{\boldsymbol{\mu}}_{\mathbf{Ld}}$ will have a unit norm (i.e. $\|\dot{\boldsymbol{\mu}}_{\mathbf{Ld}}\| = 1$) when $\mathbf{d}$ is a unit vector. The computation of $\gamma^t_{\mathbf{h}}$ and $\gamma^n_{\mathbf{h}}$ is thus simplified into:

$$\gamma^t_{\mathbf{Ld}} = \rho\|\mathbf{d}^T\ddot{\mathbf{A}}^t\mathbf{d}\|, \forall\mathbf{d} : \|\mathbf{d}\| = 1 \tag{17}$$

$$\gamma^n_{\mathbf{Ld}} = \rho\|\mathbf{d}^T\ddot{\mathbf{A}}^n\mathbf{d}\|, \forall\mathbf{d} : \|\mathbf{d}\| = 1 \tag{18}$$

In the proposed experimental design, the maximum of these curvature measures are used, where:

$$\gamma^t_{max} = \max_{\|\mathbf{d}\|=1}\gamma^t_{\mathbf{Ld}} \tag{19}$$

$$\gamma^n_{max} = \max_{\|\mathbf{d}\|=1}\gamma^n_{\mathbf{Ld}} \tag{20}$$

As mentioned above, in formulating the MOO for the design of experiments, two design criteria have been taken into account. The first is that the experiment should be designed to maximize the informativeness of the data for parameter estimation. In this case, we employ an information metric based on the FIM. Meanwhile, the second design criterion in the MOO aims to minimize both the parameter-effect and intrinsic curvatures. The multi-objective optimization formulation offers certain advantages that there is no need to prioritize any one of the criteria beforehand. Instead, we generate the Pareto set or Pareto frontier representing the set of solutions for which we can not improve the value of one objective function without negatively affecting the other(s) [26].

Considering the kinetic ODE model given in Eq. 3, our multi-objective formulation using the D-optimal criterion is given by:

$$\begin{aligned} &\max_{\mathbf{x_0},\mathbf{t_{sp}},\mathbf{u}(t)} \prod_i \lambda_i \\ &\min_{\mathbf{x_0},\mathbf{t_{sp}},\mathbf{u}(t)} \gamma^t_{max} + \gamma^n_{max} \end{aligned} \tag{21}$$

subject to:

$$\begin{aligned} \frac{d\mathbf{x}(\hat{\boldsymbol{\theta}},t)}{dt} &= \mathbf{g}(\mathbf{x}(\hat{\boldsymbol{\theta}},t),\mathbf{u},\hat{\boldsymbol{\theta}}) \\ \mathbf{x}(\hat{\boldsymbol{\theta}},0) &= \mathbf{x}_0 \\ \mathbf{x}_0^L \leq \mathbf{x}_0 &\leq \mathbf{x}_0^U \\ \mathbf{u}_j^L \leq \mathbf{u}_j &\leq \mathbf{u}_j^U \end{aligned} \tag{22}$$

where λ_i is the i-th eigenvalue of FIM (Eq. 5). The first objective function can be substituted with other FIM-based metrics (see Table 1). The parameter vector $\hat{\boldsymbol{\theta}}$ is either an initial guess of the parameter values or the parameter estimates from the current iteration of an iterative model identification procedure [6]. The decision variables may include the initial condition of the states $\mathbf{x}_0$, the sampling time points of measurements $\mathbf{t}_{sp}$, and the dynamic input $\mathbf{u}(t)$. In the case study below, we considered control vector parametrization (CVP) of the input $u_i(t)$ as illustrated in Fig. 2.

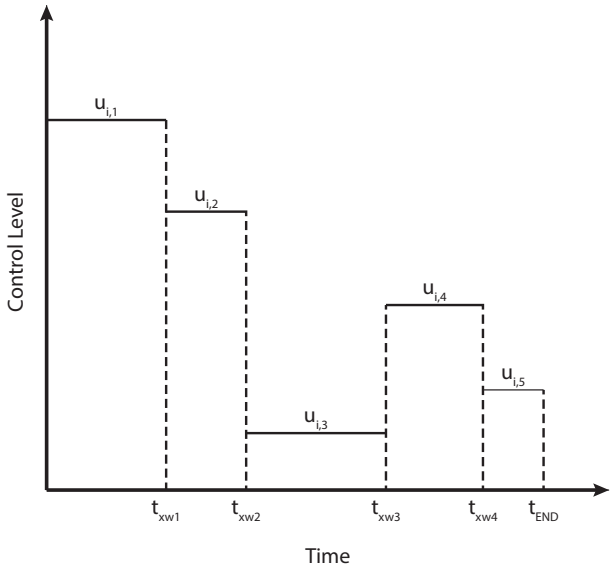

Figure 2. Control vector parametrization of input profiles. In the Baker yeast case study, we implemented piecewise constant input profiles with u_i=[$u_{i,1}$, $u_{i,2}$, $u_{i,3}$, $u_{i,4}$, $u_{i,5}$] and 4 switching times: t_{sw1}, t_{sw2}, t_{sw3}, and t_{sw4}.

Peer-reviewed version available at Processes 2017, 5, 63; doi:10.3390/pr5040063

2.2. Numerical implementation of the curvature-based MOO design

As described in the previous section, the parameter-effect and intrinsic curvatures require the computation of the first and second-order model sensitivities. For the ODE model in Eq. 3, the first-order sensitivities can be calculated according to:

$$\hat{\mathbf{F}} = \dot{\mathbf{F}}(\hat{\boldsymbol{\theta}}; \mathbf{x}) = \frac{\partial \mathbf{F}(\mathbf{x}(t, \mathbf{u}, \boldsymbol{\theta}))}{\partial \mathbf{x}} \frac{\partial \mathbf{x}(t, \mathbf{u}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} / \boldsymbol{\theta}} \Big|_{\hat{\boldsymbol{\theta}}} \tag{23}$$

The sensitivities in the above equation are normalized with respect to the parameter values. The last term on the right hand side is the first-order sensitivities of the ODE model, which obey the following differential equation:

$$\frac{d}{dt} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}} = \frac{\partial \mathbf{g}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}} + \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}, \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}} \Big|_{t=0} = 0 \tag{24}$$

Here, we have assumed that $\mathbf{x}_0$ is not part of the parameter estimation, but such assumption can be easily relaxed. In the case study, the sensitivities $\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}}$ were computed by solving the ODE in Eq. 24 simultaneously with Eq. 3, following a procedure known as the direct differential method [27]. Meanwhile, the Hessian matrix was approximated using a finite difference method, as follows:

$$\hat{\mathbf{F}}_{ijk} = \begin{cases} \frac{F_i(\boldsymbol{\theta}+\Delta\theta_j \mathbf{e}_j) - 2F_i(\boldsymbol{\theta}) + F_i(\boldsymbol{\theta}-\Delta\theta_j \mathbf{e}_j)}{\Delta\theta_j^2 / \theta_j^2}, & \text{for } j = k \\ \frac{F_i(\boldsymbol{\theta}+\Delta\theta_j \mathbf{e}_j+\Delta\theta_k \mathbf{e}_k) - F_i(\boldsymbol{\theta}+\Delta\theta_j \mathbf{e}_j-\Delta\theta_k \mathbf{e}_k) - F_i(\boldsymbol{\theta}-\Delta\theta_j \mathbf{e}_j+\Delta\theta_k \mathbf{e}_k) + F_i(\boldsymbol{\theta}-\Delta\theta_j \mathbf{e}_j-\Delta\theta_k \mathbf{e}_k)}{(\Delta\theta_j / \theta_j)(\Delta\theta_k / \theta_k)}, & \text{for } j \neq k \end{cases} \tag{25}$$

where $\mathbf{e}_j$ is the j-th elementary vector and using 1% parameter perturbations (i.e. $\Delta\theta_j / \theta_j$=0.01). The second order sensitivities above are also normalized with respect to the parameter values.

Meanwhile, the curvature measures γ^t_{max} and γ^n_{max} in Eqs. 19 and 20 were calculated from the Hessian matrix using the alternating least squares (ALS) method [28], an algorithm created to find the maximum singular value σ_{max} of a three-dimensional matrix. Based on the definitions in Eqs. 19 and 20, the maximum curvature measures can be determined by computing the maximum singular values of the matrices $\rho\ddot{\mathbf{A}}^t$ and $\rho\ddot{\mathbf{A}}^n$, respectively. More specifically, we implemented the ALS method to solve for:

$$\sigma_{max}(\mathbf{B}) = \max_{\|\mathbf{r}\|=\|\mathbf{s}\|=1} \sum_{i=1}^{m} \mathbf{r_i}\, \mathbf{s}^T \mathbf{B_i} \mathbf{s} \tag{26}$$

where $\mathbf{B}$ is either $\rho\ddot{\mathbf{A}}^t$ and $\rho\ddot{\mathbf{A}}^n$. The ALS algorithm started with initial guess values of the vectors $\mathbf{r}$ and $\mathbf{s}$, and used the above equation to solve for one variable while fixing the other in an alternating manner. Zhang and Golub showed that the method linearly converges in a neighbourhood of the optimal solution [28].

In the case study, the MOO problem was solved using the non-dominated sorting genetic algorithm II (NSGAII) in MATLAB, producing a Pareto frontier in the space of the objective functions [29]. We employed a population size of 300 and set the number of generations to 50 times the number of parameters (i.e. 1450). We recasted a maximization of an objective function as the minimization of its negative counterpart. The optimal design was selected from the Pareto frontier by balancing the trade-offs among the objective functions. More specifically, we first normalized the objective functions such that their values on the Pareto frontier ranged between 0 and 1. Finally, we chose among the solutions on the Pareto frontier, the one which minimizes the Euclidean distance of all (normalized) objective functions as the final design.

3. Results

3.1. MBDOEs of Baker yeast fermentation model

We evaluated the performance of the proposed MBDOE in an application to a kinetic model of fed-batch fermentation of Baker's yeast [21,22]. In addition to D-optimal criterion, we also

implemented A-optimal, E-optimal and modified E-optimal criteria (see Table 1) with our MOO MBDOE. We compared the performance of our method to other MBDOEs, including (a) FIM-based MBDOEs, i.e. D-optimal, A-optimal, E-optimal and modified E-optimal designs; (b) D-optimal design with a curvature threshold [22]; (c) Q-optimal MBDOE [20]; and (d) MOO MBDOE using parameter correlation [17]. In total, we applied and compared 14 MBDOE methods. For the optimizations in (a), (b) and (c), we employed the enhanced scatter search metaheuristic (eSSm) algorithm [30–32]. For the MOO in (d), we used the optimization algorithm and optimal Pareto point selection as described in the previous section.

In the fed-batch fermenter model, cellular growth and product formation are captured by the biomass variable x_1, which is assumed to rely on a single substrate variable x_2. The fermenter operates at a constant temperature and the feed is free from product. The model equations are given by:

$$\begin{aligned} \frac{dx_1}{dt} &= (r - u_1 - \theta_4)x_1,\ x_1(0) = x_{10} \\ \frac{dx_2}{dt} &= -\frac{rx_1}{\theta_3} + u_1(u_2 - x_2),\ x_2(0) = 0.01 \\ r &= \frac{\theta_1 x_2}{\theta_2 + x_2} \end{aligned} \tag{27}$$

where the input u_1 is the dilution factor (in the range of 0.05 to 0.20 hour^{-1}) and the input u_2 is the substrate concentration in the feed (in the range of 5 to 35 g/L). In the model, the biomass growth follows a Monod-type kinetics. The parameters θ_1 and θ_2 are the Monod kinetic parameters, θ_3 is the yield coefficient, and θ_4 is the cell death rate constant.

In the MBDOE, the design variables consisted of the initial condition of the biomass $x_1(0)$ in the range between 1 and 10 g/L, 10 measurement sampling times (t_{sp}), and the inputs $u_1(t)$ and $u_2(t)$. The piecewise-constant dynamic inputs were each parametrized using the CVP as shown in Fig. 2. Thus, the MOO was performed with 29 design parameters ($x_1(0)$, 10 t_{sp}'s, 10 $u_{i,j}$'s, and 8 t_{sw}'s). The length of the time interval between two successive measurement sampling points was constrained to between 1 and 20 hours, while that between two input switching times was bounded between 2 and 20 hours. The calculations of the Jacobian and Hessian matrices in MBDOEs were done using parameter values $\boldsymbol{\theta}_d = [\theta_1, \theta_2, \theta_3, \theta_4]$=[0.5, 0.5, 0.5, 0.5] [17,33], which were different from the "true" parameter values used for noisy data generation in the next section. The reason for using a different parameter set in the MBDOE from the true values is to emulate the typical scenario in practice where one would start only with an estimate or guess of the model parameters. Fig. 3 and 4 and show the optimal dynamic inputs and data sampling times resulting from all MBDOE methods mentioned above (see also the Pareto frontiers in Supplementary Information). Meanwhile, Table 2 gives the optimal initial biomass concentration $x_1(0)$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2017 doi:10.20944/preprints201709.0063.v1

Peer-reviewed version available at Processes 2017, 5, 63; doi:10.3390/pr5040063

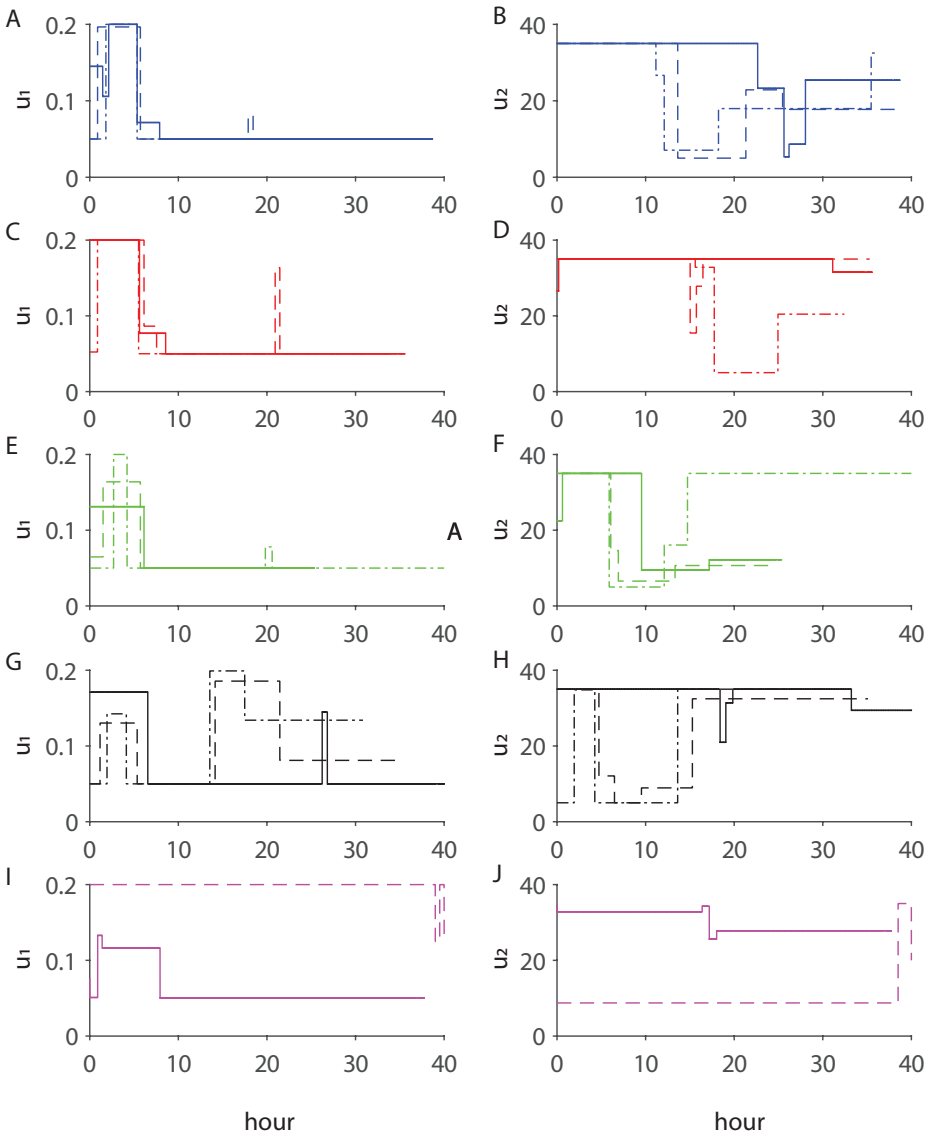

Figure 3. Optimal dilution factor and feed substrate concentration. Optimal dilution factor (u_1 in hour^{-1}, left panels) and feed substrate concentration (u_2 in g/L, right panels). (**A-B**) D-optimal (blue). (**C-D**) A-optimal (red). (**E-F**) E-optimal (green). (**G-H**) modified E-optimal (black). On panels (**A-H**), the optimal u_1 and u_2 using FIM-based criteria are shown in solid line. Those using FIM-based criteria combined with curvatures are shown in dashed line, while those using FIM-based criteria combined with parameter correlation are drawn in dashed-dot line. (**I-J**) threshold curvature (magenta, solid line), and Q-optimal design (magenta, dashed line).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2017 doi:10.20944/preprints201709.0063.v1

Peer-reviewed version available at Processes 2017, 5, 63; doi:10.3390/pr5040063

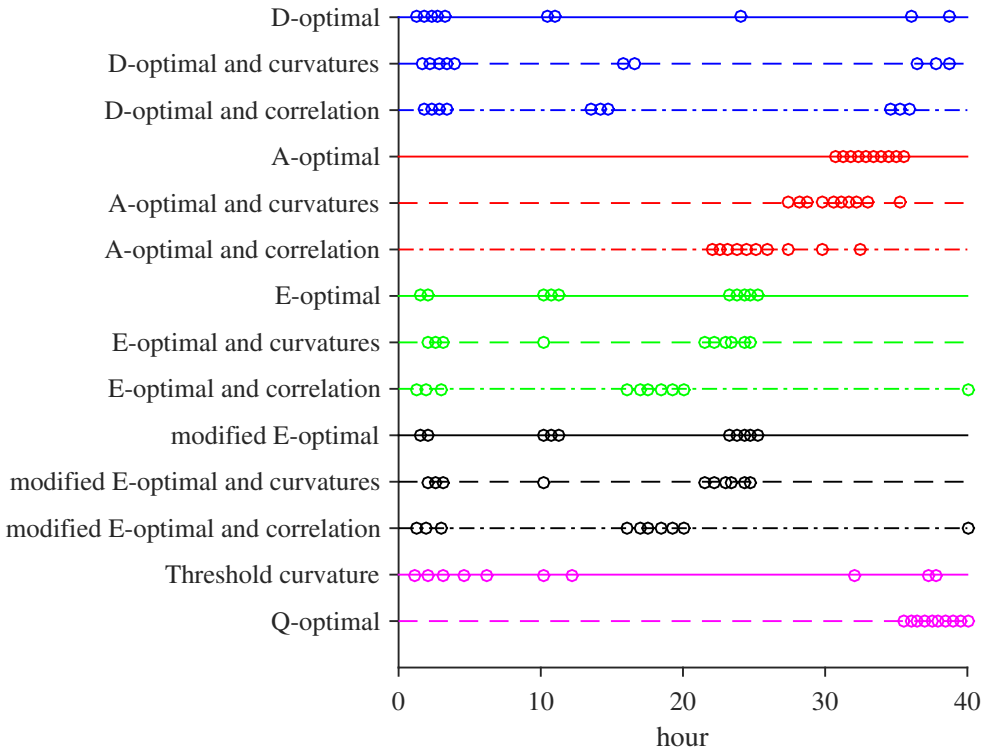

Figure 4. Optimal sampling grid from MBDOEs. Simple FIM-based criteria in continuous line, FIM-based criteria combined with curvatures in dashed line, FIM-based criteria combined with parameter correlation in dashed-pointed line. Dots indicate the sampling times.

Table 2. Optimal initial condition of biomass $x_1(0)$ (g/L) from MBDOEs.

Design criterion	$x_1(0)$
D-optimal	10.0
MOO D-optimal and curvatures	10.0
MOO D-optimal and correlation	10.0
A-optimal	10.0
MOO A-optimal and curvatures	9.9
MOO A-optimal and correlation	10.0
E-optimal	10.0
MOO E-optimal and curvatures	10.0
MOO E-optimal and correlation	10.0
modified E-optimal	10.0
MOO modified E-optimal and curvatures	10.0
MOO modified E-optimal and correlation	10.0
threshold curvature	8.2
Q-optimal	5.5

3.2. Performance evaluation

For each of the optimal experimental designs above, we generated *in silico* datasets by simulating the ODE model using the parameter values $\boldsymbol{\theta}^*$=[0.31, 0.18, 0.55, 0.05], as reported in previous publications [17,33]. We subsequently added independent and identically distributed (i.i.d.) Gaussian random white noise to the model simulations using a variance σ^2 of 0.04 for both $x_1(t)$ and

$x_2(t)$ [17,33]. For each *in silico* dataset, we then performed a parameter estimation using the resulting data (y_1 and y_2) by maximum likelihood estimation, i.e. by minimizing:

$$\Phi(\theta) = \frac{1}{\sigma^2}\sum_{i=1}^{10}[y_1(t_i) - x_1(t_i, \boldsymbol{\theta})]^2 + [y_2(t_i) - x_2(t_i, \boldsymbol{\theta})]^2 \tag{28}$$

We also employed the following constraints for $\boldsymbol{\theta}$ in the optimization above:

$$0.05 \leq \theta_1, \theta_2, \theta_3 \leq 0.98$$

and

$$0.01 \leq \theta_4 \leq 0.98.$$

Finding the globally optimal solution to the parameter estimation in Eq. (4) is challenging. Here, we solved the constrained parameter optimization problem using the interior-point algorithm (implemented by the subroutine *fmincon* function in MATLAB) with the true parameter values $\boldsymbol{\theta}^*$ as the initial guess. By employing the true values as the initial starting point of the optimization, we expect that the parameter accuracy would mainly be affected by the experimental design and not by the ability of the parameter optimization algorithm to find the globally optimal solution.

We repeated the in silico data generation and parameter estimation as described above for 100 times, which resulted in a set of 100 parameter estimates. The performance of each MBDOE was assessed by the average accuracy of the parameter estimates, measured by the average of the normalized Mean Squared Error (nMSE):

$$\overline{\text{nMSE}} = \frac{1}{4}\sum_{i=1}^{4}\text{nMSE}_\text{i} \tag{29}$$

where:

$$\text{nMSE}_\text{i} = \frac{\text{variance}(\hat{\theta}_i) - \text{bias}^2(\hat{\theta}_i)}{(\theta_i^*)^2},\ i = 1, 2, 3, 4 \tag{30}$$

The variance of $\hat{\theta}_i$ was computed using the set of 100 parameter estimates, while the bias was calculated as the difference between the average of $\hat{\theta}_i$ and θ_i^*. Table 3 gives the average nMSE of the parameter estimates from each MBDOE under consideration.

Table 3. MBDOE performance on the fed-batch fermentation of Baker's yeast model. The overall parameter accuracy is represented by the average of nMSE. The reported parameter values and errors are the averages and standard deviations from 100 repeated runs of parameter estimation.

Design criterion	$\overline{\text{nMSE}}$	$\theta_1 \pm SD_{\theta_1}$	$\theta_2 \pm SD_{\theta_2}$	$\theta_3 \pm SD_{\theta_3}$	$\theta_4 \pm SD_{\theta_4}$
D-optimal	7.06×10^{-3}	0.3107 ± 0.0102	0.1831 ± 0.0276	0.5505 ± 0.0125	0.0502 ± 0.00
MOO D-optimal and curvatures	4.71×10^{-3}	0.3099 ± 0.0056	0.1825 ± 0.0233	0.5496 ± 0.0099	0.0499 ± 0.00
MOO D-optimal and correlation	5.36×10^{-3}	0.3117 ± 0.0134	0.1781 ± 0.0151	0.5543 ± 0.0270	0.0508 ± 0.00
A-optimal	2.35×10^{-1}	0.3294 ± 0.0659	0.2399 ± 0.1387	0.5841 ± 0.1083	0.0558 ± 0.01
MOO A-optimal and curvatures	1.42	0.3669 ± 0.0947	0.5267 ± 0.2230	0.5548 ± 0.1333	0.0510 ± 0.02
MOO A-optimal and correlation	4.82	0.0863 ± 0.0499	0.8927 ± 0.2555	0.2879 ± 0.1928	0.0177 ± 0.02
E-optimal	8.01×10^{-2}	0.3180 ± 0.0420	0.2026 ± 0.0956	0.5473 ± 0.0159	0.0496 ± 0.00
MOO E-optimal and curvatures	3.33×10^{-3}	0.3083 ± 0.0095	0.1829 ± 0.0164	0.5502 ± 0.0183	0.0500 ± 0.00
MOO E-optimal and correlation	8.19×10^{-3}	0.3108 ± 0.0164	0.1824 ± 0.0213	0.5552 ± 0.0304	0.0509 ± 0.00
modified E-optimal	6.99×10^{-2}	0.3137 ± 0.0165	0.1986 ± 0.0920	0.5498 ± 0.0144	0.0502 ± 0.00
MOO modified E-optimal and curvatures	3.44×10^{-4}	0.3095 ± 0.0036	0.1789 ± 0.0034	0.5491 ± 0.0073	0.0500 ± 0.00
MOO modified E-optimal and correlation	2.27×10^{-3}	0.3088 ± 0.0048	0.1820 ± 0.0160	0.5486 ± 0.0047	0.0496 ± 0.00
threshold curvature	1.29×10^{-2}	0.3144 ± 0.0307	0.1857 ± 0.0339	0.5500 ± 0.0155	0.0502 ± 0.00
Q-optimal	1.91×10^{-2}	0.3085 ± 0.0178	0.1757 ± 0.0216	0.5514 ± 0.0236	0.0504 ± 0.01

4. Discussion

As shown in Fig. 3, the MBDOEs prescribed manipulating the input $u_1(t)$ mostly at the beginning of the experiment and the input $u_2(t)$ for the entire duration of the experiment. For the majority of the MBDOEs in this study, the optimal sampling times spread unevenly over the duration of the experiment (see Fig. 4). A more detailed comparison between Fig. 3 and 4 showed that the optimal sampling points were typically placed before and after a change in the dynamic inputs $u_1(t)$ and $u_2(t)$. The exception to this observation was for the optimal design using A-optimal criterion, which gave the worst parameter accuracy among the MBDOEs considered.

The consideration of model curvature using the proposed MOO MBDOE generally led to improved parameter accuracy over using only model curvature (i.e. Q-optimal and threshold curvature) or using only FIM-based criteria. The lowest nMSE came from the MOO MBDOE design using modified E-optimal with model curvature. In comparison to MOO MBDOE using parameter correlation, employing model curvature in the MOO framework gave better experimental designs with lower average nMSEs. Meanwhile, Q-optimality and curvature thresholding strategies provided better nMSEs than the majority of the FIM-based criteria, except the D-optimal design. Finally, the optimal experiments based on A-optimal criterion, either alone or in MOO MBDOE, performed poorly. The poor performance of A-optimal design has also been reported in a previous publication [17].

The obvious drawback of curvature-based MBDOEs in comparison to FIM-based strategies is the higher computational cost associated with computing the Hessian matrix. While the number of first-order sensitivities (Jacobian) increases linearly with the number of parameters p, the number of second-order sensitivities scales with p^2. Fortunately, the calculation of the Hessian matrix can be easily parallelized and implemented using multiple computing cores. In practice, one often focuses on only a subset of the model parameters, and therefore the MBDOE is typically done for a handful of parameters.

5. Conclusions

Existing MBDOE methods mostly rely on the FIM to define information criteria. Since the FIM is based on first-order sensitivities with respect to the model parameters, the related MBDOEs may perform poorly for highly nonlinear models. Here, a new model-based design of experiment using multi-objective optimization framework was presented, employing the maximization of a FIM-based

information metric and the minimization of model curvatures. The application to a model of the fermentation of Baker's yeast demonstrated that accounting model nonlinearity through model curvatures in designing experiment could lead to improved parameter accuracy over using only a FIM-based criterion. The proposed MOO MBDOE also outperformed other curvature-based designs, including the Q-optimality and curvature thresholding, and another MOO MBDOE strategy using parameter correlation. The use of the MOO framework further gives flexibility to accommodate other criteria that may arise in a particular application, in the design of experiments.

Supplementary Materials: The following are available online at www.mdpi.com/link, Supplementary Information.

Acknowledgments: The authors would like to acknowledge funding from ETH Zurich and the Ministry of Education of Singapore.

Author Contributions: R.G. conceived and designed the study; E.M. and S.S. performed the design of experiments and parameter estimations; E.M., S.S. and R.G. analyzed the data; E.M., S.S. and R.G. wrote the paper.

The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

MBDOE	Model-based design of experiments
FIM	Fisher Information Matrix
MOO	Multi-Objective Optimization
RMS	Root Mean Square
ODE	Ordinary Differential Equation
MLE	Maximum Likelihood Estimator
CVP	Control Vector Parametrization
nMSE	normalized Mean Squared Error

Bibliography

1. Srinath, S; Gunawan, R. Parameter identifiability of power-law biochemical system models. *J. Biotechnol.* **2010**, *149*, 132-140.
2. Jia, G; Stephanopoulos, G.; Gunawan, R. Ensemble kinetic modeling of metabolic networks from dynamic metabolic profiles. *Metabolites* **2012**, *2*, 891-912.
3. Gábor, A.; Hangos, K.; Banga, J.; Szederkàyi G. Reaction network realizations of rational biochemical systems and their structural properties. *J. Math. Chem.* **2015**, *53*, 1657-1686.
4. Liu, Y.; Manesso, E.; Gunawan, R. REDEMPTION: Reduced dimension ensemble modeling and parameter estimation. *Bioinformatics* **2015**, *31*, 3387-3389.
5. Villaverde, A.F.; Banga, J.R. Structural properties of dynamic systems biology models: Idenfiability, reachability and initial conditions. *Processes* **2017**, *5*, 29.
6. Franceschini, G.; Macchietto, S. Model-based design of experiments for parameter precision: State of the art. *Chem. Eng. Sci.* **2008**, *63*, 4846-4872.
7. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2006.
8. Emery, A.F.; Nenarokomov, A.V. Optimal experiment design. *Meas. Sci. Technol.* **1998**, *9*, 864-876.
9. Franceschini, G.; Macchietto, S. Validation of a model for biodiesel production through model-based experiment design. *Ind. & Eng. Chem. Res.* **2007**, *46*, 220-232.
10. Franceschini, G.; Macchietto, S. Novel anticorrelation criteria for design of experiments: Algorithm and application. *AIChE J.* **2008**, *54*, 3221-3238.
11. Gadkar, K.G.; Gunawan, R.; Doyle III, F.J. Iterative approach to model identification of biological networks. *BMC Bioinform.* **2005**, *6*, 155.

12. Van Derlinden, E.; Bernaerts, K.; Van Impe, J.F. Simultaneous versus sequential optimal experiment design for the identification of multi-parameter microbial growth kinetics as a function of temperature. *J. theor. Biol.* **2010**, *264*, 347-355.
13. Walter, E.; Pronzato, L. Qualitative and quantitative experiment design for phenomenological models: a survey. *Automatica* **1990**, *26*, 195-213.
14. Sinkoe, A.; Hahn, J. Optimal experimental design for parameter estimation of an IL-6 signaling model. *Processes* **2017**, *5*, 49.
15. Balsa-Canto, E.; Alonso, A.A.; Banga, J.R. Computational procedures for optimal experimental design in biological systems. *IET Syst. Biol.* **2008**, *2*, 163-172.
16. Faller, D.; Klingmüller, U.; Timmer, J. Simulation methods for optimal experimental design in systems biology. *Simulation* **2003**, *79*, 717-725.
17. Maheshwari, V.; Rangaiah, G.P.; Samavedham, L. A multi-objective framework for model based design of experiments to improve parameter precision and minimize parameter correlation. *Ind. & Eng. Chem. Res.* **2013**, *52*, 8289-8304.
18. Cochran, W.G. Experiments for Nonlinear Functions. *J. Am. Stat. Assoc.* **1973**, *68*, 771-781.
19. Bates, D.M.; Watts, D.G. Relative Curvature Measures of Nonlinearity. *J. R. Stat. Soc. Ser. B* **1980**, *42*, 1-25.
20. Hamilton, D.C.; Watts, D.G. A quadratic design criterion for precise estimation in nonlinear regression models. *Technometrics* **1985**, *27*, 241-250.
21. Asprey, S.P.; Macchietto, S. Statistical tools for optimal dynamic model building. *Comput. & Chem. Eng.* **2000**, *24*, 1261-1267.
22. Benabbas, L.; Asprey, S.P.; Macchietto, S. Curvature-based methods for designing optimally informative experiments in multiresponse nonlinear dynamic situations. *Ind. & Eng. Chem. Res.* **2005**, *44*, 7120-7131.
23. Seber, G.A.F.; Wild, C.J.*Nonlinear regression*; John Wiley & Sons: Hoboken, NJ, USA, 2003.
24. Merlé, Y.; Tod, M. Impact of pharmacokinetic-pharmacodynamic model linearization on the accuracy of population information matrix and optimal design. *J. Pharmacokinet. Pharmacodyn.* **2001**, *28*, 363-388.
25. Bogacka, B.; Wright, F. Comparison of two design optimality criteria applied to a nonlinear model. *J. Biopharm. Stat.* **2004**, *14*, 909-930.
26. Rangaiah, G.P. *Multi-objective optimization: techniques and applications in chemical engineering*; Vol. 1; World Scientific: Singapore, 2008.
27. Varma, A.; Morbidelli, M.; Wu, H. *Parametric Sensitivity in Chemical Systems*; Cambridge University Press: Cambridge, UK, 1999.
28. Zhang, T.; Golub, G.H. Rank-One Approximation to High Order Tensors. *SIAM J. Matrix Anal. & Appl.* **2001**, *23*, 534-550.
29. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182-197.
30. Egea, J.A.; Martí, R.; Banga, J.R. An evolutionary method for complex-process optimization. *Comput. & Oper. Res.* **2010**, *37*, 315-324.
31. Egea, J.A.; Rodriguez-Fernandez, M.; Banga, J.R.; Martí, R. Scatter search for chemical and bioprocess optimization. *J. Global Optim.* **2007**, *37*, 481-503.
32. Rodriguez-Fernandez, M.; Egea, J.A.; Banga, J.R. Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinformatics* **2006**, *7*, 483.
33. Zhang, Y.; Edgar, T.F. PCA combined model-based design of experiments (doe) criteria for differential and algebraic system parameter estimation. *Ind. & Eng. Chem. Res.* **2008**, *47*, 7772-7783.