Using Prior Parameter Knowledge in Model-Based Design of Experiments for Pharmaceutical Production

Ali Shahmohammadi[1] and Kimberley McAuley[2]

[1]University of Texas at Austin
[2]Queens University

May 5, 2020

Abstract

Sequential model-based design of experiments (MBDoE) uses information from previous experiments to select run conditions for new experiments. Computation of the objective functions for popular MBDoE can be impossible due to a non-invertible Fisher Information Matrix (FIM). Previously, we evaluated a leave-out (LO) approach that design experiments by removing problematic model parameters from the design process. However, the LO approach can be computationally expensive due to its iterative nature and some model parameters are ignored. In this study, we propose a simple Bayesian approach that makes the FIM invertible by accounting for prior parameter information. We compare the proposed Bayesian approach to the LO approach for designing sequential A-optimal experiments. Results from a pharmaceutical case study show that the Bayesian approach is superior, on average, to the LO approach for design of experiments. However, for subsequent parameter estimation, a subset-selection-based LO approach gives better parameter values than the Bayesian approach.

Introduction

Mathematical models are used in chemical and pharmaceutical industries for analysis, design and control of chemical processes and for maximizing product quality and profit.[1,2] Especially in pharmaceutical industries, models are important for Quality by Design and development of continuous manufacturing processes, which are becoming more widespread.[3–5] Mathematical models for pharmaceutical product development can be either empirical or mechanistic.[5–7] Although empirical models are commonly used for pharmaceutical processes, they cannot reliably predict the system behavior outside the range of operating conditions used for model development.[8] Therefore, fundamental models, based on underlying chemistry and physics, are preferred.[9] These models usually contain unknown parameters that require estimation using experimental data.[10] To obtain informative data, it is advantageous to carefully plan the experiments aimed at parameter estimation using design of experiment (DoE) techniques.[11] As shown in Table 1, optimal model-based design-of-experiments (MBDoE) techniques select experiments to minimize uncertainties in parameters estimates or model predictions.[12–14] MBDoE techniques are effective because they account for the structure of the model as well as parameter and measurement uncertainties when selecting new run conditions.[13,15] Other benefits of MBDoE techniques, compared to traditional factorial designs, are that they can be readily used to design any number of experiments, e.g., one, three or seven experiments, depending on available resources for experimentation.[15,16] MBDoE techniques have been developed to satisfy a variety of objectives including minimizing total variances of parameter estimates, minimizing the average variance of model predictions, and designing experiments for model discrimination.[16]

Table 1 shows several MBDoE objective functions that have been used for development of chemical and pharmaceutical production models.[15–17] If modelers are interested in obtaining accurate parameter estimates

for their model, A-, D- or E- optimal designs can be selected.[15–17] Alternatively, G- and V-optimal designs focus on obtaining accurate model predictions at specified operating conditions of interest to the modeler.[18–21] All of these MBDoE techniques in Table 1 require computation of the inverse of the Fisher Information Matrix (**FIM**) when selecting experimental settings.[21–23] The **FIM** carries information about how changes in parameter values can affect the model predictions and is therefore crucial for both MBDoE calculations and parameter inference.[24] For nonlinear models, which are common in chemical and pharmaceutical applications, computation of the**FIM** requires linearizing the model around some nominal parameter values.[17,25] If these nominal parameter values are significantly different from the corresponding true values, the selected MBDoE settings may lead to experimental data that are not very informative.[17,25] Sequential design approaches are appealing because they enable updating of the parameter values, as well as the experimental strategy, as more data become available.[26] Using sequential experimental designs, valuable information from old experimental data can be used, which might have been collected for other objectives than model development.[27,28]

Computation of the objective functions for sequential MBDoE is problematic if the **FIM** is noninvertible or ill-conditioned. Typical causes are limited experimental data, strongly correlated influences of different parameters, and parameters with little or no influence on the model predictions.[29] In chemical, biochemical and pharmacological systems, models often contain a large number of kinetic and transport parameters (e.g., 10-80 parameters) which may result in noninvertible/ill-conditioned**FIM** s.[30–34] To avoid this problem, several approaches have been considered during sequential MBDoE calculations including parameter subset selection,[14,29,35]pseudoinverse methods,[21,36] Tikhonov regularization,[37–40] and Bayesian approaches.[13,41,42]

The parameter-subset-selection approach uses a model-reduction perspective.[35,43,44] In one methodology, parameters are ranked from most-estimable to least-estimable so that problematic (low-ranked) parameters can be recognized and fixed at their nominal values.[35,45] In this way, experiments can be designed using a well-conditioned reduced **FIM** that ignores problematic parameters. Alternatively, pseudoinverse methods approximate the inverse of the **FIM** (e.g., using the Moore-Penrose pseudoinverse) during MBDoE calculations.[21,36,46] In Tikhonov regularization, a penalty is added to diagonal elements of the**FIM** to make it invertible.[29,38–40] Bayesian MBDoE using linear models results in Tikhonov penalties that account for prior knowledge about parameters. However, for nonlinear models, the situation can be considerably more complex, depending on how the nonlinearity is treated.[13,41,42] There is little information in the literature regarding which approach is most effective. In two previous articles, we considered pharmaceutical case studies involving noninvertible **FIM** s. Two different approaches were compared: i) a subset-selection-based approach that leaves out problematic parameters (LO approach) and ii) a simpler approach that uses a Moore-Penrose pseudoinverse in place of$\mathbf{FIM}^{-1}$(PI approach).[21,46] These case studies suggest that the LO approach is often superior to the PI approach for designing both A- and V-optimal experiments.[21,46] A shortcoming of the LO approach is that it can be complicated and computationally expensive due to changes in the subset of parameters that is left out during MBDoE calculations. This complication motivates us to find a more convenient approach to deal with singular **FIMs** during MBDoE.

The focus of the current study is on a simplified Bayesian approach for dealing with singular/ill-conditioned **FIMs** during MBDoE. Bayesian approaches have been used in several past MBDoE studies for chemical and biochemical systems.[47–49] The main benefit of the Bayesian MBDoE framework is that it accounts for prior knowledge about plausible values of the model parameters.[13] However, many researchers raise concerns about the use of Bayesian approaches in practical engineering systems.[13,50] Disadvantages of the Bayesian approach include uncertainty about the reliability of assumptions made when specifying prior information.[49–51] Undesirable computational complexity can also arise, depending on the assumptions that are made. As a result, Bayesian MBDoE has not enjoyed widespread applications in chemical process modeling.

Table 1. Optimality criteria for model-based design of experiments [21,46]

Optimality Criterion	**Description**
$J_A = trace\left((\mathrm{FIM})^{-1}\right)$	**A-optimal** design minimizes total parameter variance.

Optimality Criterion	Description
$J_D = det\left((\mathrm{FIM})^{-1}\right)$	**D-optimal** design minimizes the volume of the joint confidence interval for the pa
$J_E = \lambda_{\max}(\mathrm{FIM}^{-1})$	**E-optimal** design minimizes the largest eigenvalue of the **FIM**, thereby minimizin
$J_G = max\ \left(\mathrm{diag}\left(\mathbf{W}\left(\mathrm{FIM}\right)^{-1}\mathbf{W^T}\right)\right)$	**G-optimal** design minimizes the maximum variance of model predictions at user-s
$J_V = trace\left(\mathbf{W}\left(\mathrm{FIM}\right)^{-1}\mathbf{W^T}\right)$	**V-optimal** design minimizes the total variance of model predictions at user-specifi

The objective of the current article is to formulate and test a simple Bayesian MBDoE approach that is readily usable by model developers. The effectiveness of the proposed Bayesian approach is compared to that of the LO approach for designing A-optimal experiments when the**FIM** is noninvertible. We use the pharmaceutical case study of Domagalski et al., (2015), which is of interest to our industrial sponsor.[6,21] The associated dynamic model uses Michaelis-Menten kinetics and enzyme-catalyzed reactions to describe the production of a pharmaceutical agent.[52] The remainder of this article is organized as follows. First, background on the**FIM** and sequential A-optimal design is presented. Next, details of the Bayesian and LO approaches for parameter estimation and experimental design are presented. A simple Bayesian approach is proposed and a pharmaceutical case study is presented. Results obtained using Monte Carlo (MC) simulations are provided, revealing that the proposed Bayesian approach is superior to the LO approach for this case study.

Background Information

Fisher Information Matrix for Nonlinear Models

Consider the nonlinear model:

$\mathbf{Y} = \mathbf{g}\left(\mathbf{d}, \theta\right) + \varepsilon (1)$

where $\mathbf{Y} \in \mathbf{R}^N$ is a vector of stacked measured responses, $\mathbf{g}$ is the solution of equations that describe the system, $\mathbf{d} \in \mathbf{R}^{r\times D}$ is a matrix of experimental settings (for r runs with D decision variables specified for each), $\theta \in \mathbf{R}^p$ is the vector of model parameters and$\varepsilon \in \mathbf{R}^N$ is a vector of a measurement noise with diagonal covariance matrix$\mathbf{\Sigma_y} \in \mathbf{R}^{N\times N}$. For dynamic multi-response models with n sample times per run and $v$response variables, the total number of data values is $N = nvr$. The**FIM** is computed using a parametric sensitivity matrix$\mathbf{S} \in \mathbf{R}^{N\times p}$ with elements:

$S_{\mathrm{ij}} = \left.\frac{\partial g(\mathbf{d},\theta)}{\partial \theta_j}\right|_{\hat{\theta}_{k\neq j}} (2)$

computed by linearizing the model around the best currently-available parameter values:[53]

The elements of $\mathbf{S}$ should be scaled using parameter uncertainties s_{θ_j} and measurement uncertaintiess_{y_i} to reflect the modeler's prior knowledge:[54]

$Z_{\mathrm{ij}} = S_{\mathrm{ij}}\frac{s_{\theta_j}}{s_{y_i}}$ (3)

resulting in a scaled sensitivity matrix $\mathbf{Z}$. The **FIM** is related to $\mathbf{Z}$ by:

$\mathbf{FIM} = \mathbf{Z}^T\mathbf{Z}$ (4)

When performing sequential MBDoE calculations, $\mathbf{Z}$ contains two parts:[21,46]

$\mathbf{Z} = \begin{bmatrix}\mathbf{Z}_{\mathrm{old}} \\ \mathbf{Z}_{\mathrm{new}}\end{bmatrix}$ (5)

where $\mathbf{Z}_{\text{old}}$ corresponds to experimental settings and data from old experiments. The elements of$\mathbf{Z}_{\text{old}}$ are fixed during sequential MBDoE and elements of $\mathbf{Z}_{\text{new}}$ are determined by the optimizer. After each sequential design, elements of$\mathbf{Z}_{\text{old}}$ are updated based on the new parameter values and the number of rows in$\mathbf{Z}_{\text{old}}$ increases due to the recent experiments.

Parameter estimation with a noninvertible FIM

When estimating parameters, the **FIM** should be invertible, otherwise unique estimates for the parameters cannot be obtained.[22,29] Several regularization approaches have been used to overcome this problem.[38,39,55] One popular approach is to estimate a subset of the model parameters that are estimable, with the remaining parameters fixed at nominal values.[23,45,56] Table 2 shows computational steps for a commonly used orthogonalization-based approach that ranks parameters from the most-estimable so problematic (unranked) parameters that lead to a noninvertible **FIM** can be determined.[45,54] The ranking starts by computing the magnitude of each column of the scaled sensitivity matrix $\mathbf{Z}$(Step 1). The parameter corresponding to the column with the highest magnitude is selected as the most-estimable parameter (Step 2). The columns of $\mathbf{Z}$ are then regressed onto columns of$\mathbf{X}_k$, a matrix that contains columns from $\mathbf{Z}$ that correspond to the ranked parameters (Step 3). Residual matrix$\mathbf{R}_k$ is then computed to remove correlation between columns for the unranked parameters and columns for the parameters that have already been ranked (Step 4). The next-most-estimable parameter is the one with the largest magnitude among columns of$\mathbf{R_k}$. In Step 5, the column corresponding to the next-most-estimable parameter is selected from the original$\mathbf{Z}$ matrix and included in $\mathbf{X}_k$, resulting in matrix $\mathbf{X}_{k+1}$. Steps two to five are repeated to produce a ranked list with up to p parameters. The ranking stops when all of the parameters are ranked or at the iteration where$\mathbf{X}_k^T\mathbf{X}_k$ (the reduced **FIM**) becomes noninvertible. The remaining unranked parameters are categorized as problematic. They either have very little influence on the predicted responses or highly correlated effects with parameters on the ranked list.[45,57] Using this orthogonalization-based ranking approach prior to parameter estimation helps to avoid numerical problems that would arise due to a noninvertible **FIM** .

Table 2. Orthogonalization algorithm [45,54]

Compute the magnitude (i.e., the Euclidean norm) of each column in the $\mathbf{Z}$ matrix. Select the column with the largest mag

and calculate the residual matrix:

$$\mathbf{R}_k = \mathbf{Z} - \hat{\mathbf{Z}}_k \quad (2.2)$$

Michaelis-Menten Case Study

Reaction scheme and dynamic model

The case study considered in the current article uses a nonlinear kinetic model based on a Michaelis-Menten batch reaction for the production of a pharmaceutical agent. Domagalski et al. (2015) used this case study to develop empirical models based on conventional DoE and response surface methodology.[6] We used the same case study to develop and test the LO approach for V-optimal MBDoE in previous work.[21] The reaction starts with reagent SM1 reacting with catalyst D and generating intermediate SM1.D via reversible reaction (1) in Figure 1. Next, intermediate SM1.D reacts with reagent SM2 to make the product P and release the catalyst (i.e., reaction (2)). There is also a possibility of generating several impurities: SM2 can react with P to generate impurity I1, SM1 can be hydrolyzed to form impurity I2, D can be deactivated with water to make I3, and P can degrade to generate I4. Table 3 provides a fundamental dynamic model for the Michaelis-Menten batch reaction system. Equations (3.11) to (3.14) show that the concentrations of SM1,

D, SM2, and P are measured, and these measurements have experimental errors. We assume that the water concentration C_{H2O} and the solution volume $V$are constant at 0.10 M and 1.0 L, respectively.

In the study by Domagalski et al., 3 rounds of simulated experiments were performed. In each round, they conducted 16 fractional-factorial runs + 4 center-point-runs (i.e., 20 experiments in each round and 60 overall). Table 4 shows Domagalski's center-point settings for their first round of experimentation. We assume that data for the 4 replicated center-point runs are available for initial parameter estimation and construction of $\mathbf{Z}_{\text{old}}$. Step-by-step computation of $\mathbf{Z}_{\text{old}}$ using these runs is described in the Supplementary Information. The duration of each simulated batch experiment is 6.0 h with measurements taken every 45 minutes, resulting in sampling at 9 times including the initial time$t = 0$. As a result, each run involves 36 measured values (i.e., 9 values each for y_{SM1}, y_D, y_{SM2} and y_P).

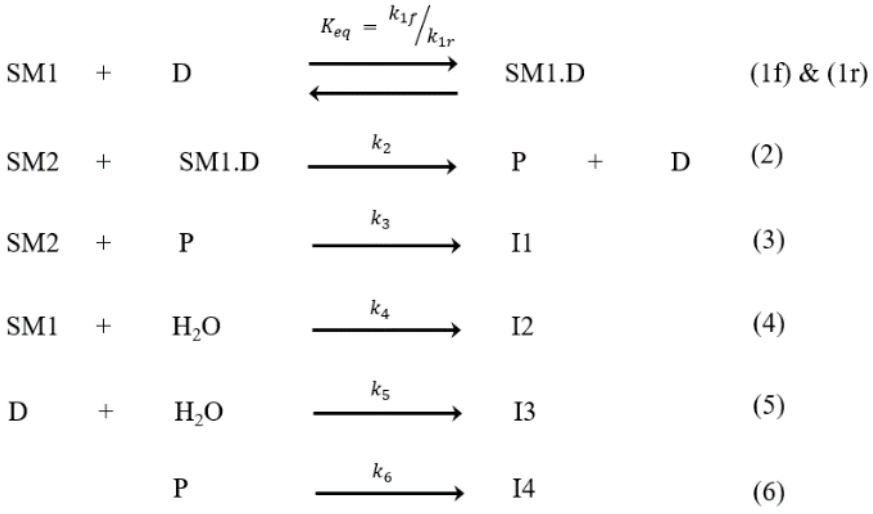

Figure . Reaction scheme for Michaelis-Menten case study [6]

Table 3. Dynamic kinetic model for the Michaelis-Menten batch reaction system[21]

$\frac{dC_{SM1}}{dt} = -k_{1f}C_{SM1}C_D + \frac{k_{1f}}{K_{\text{eq}}}C_{SM1D} - k_4C_{SM1}C_{H_2O}$	(3.1)
$\frac{dC_D}{dt} = -k_{1f}C_{SM1}C_D + \frac{k_{1f}}{K_{\text{eq}}}C_{SM1D} + k_2C_{SM2}C_{SM1D} - k_5C_DC_{H_2O}$	(3.2)
$\frac{dC_{SM1D}}{dt} = k_{1f}C_{SM1}C_D - \frac{k_{1f}}{K_{\text{eq}}}C_{SM1D} - k_2C_{SM2}C_{SM1D}$	(3.3)
$\frac{dC_{SM2}}{dt} = -k_2C_{SM2}C_{SM1D} - k_3C_{SM2}C_P$	(3.4)
$\frac{dC_P}{dt} = k_2C_{SM2}C_{SM1D} - k_3C_{SM2}C_P - k_6C_P$	(3.5)
$\frac{dC_{H_2O}}{dt} = -k_4C_{SM1}C_{H_2O} - k_5C_DC_{H_2O}$	(3.6)
$\frac{dC_{I1}}{dt} = k_3C_{SM2}C_P$	(3.7)
$\frac{dC_{I2}}{dt} = k_4C_{SM1}C_{H_2O}$	(3.8)
$\frac{dC_{I3}}{dt} = k_5C_DC_{H_2O}$	(3.9)
$\frac{dC_{I4}}{dt} = k_6C_P$	(3.10)
$y_{SM1} = C_{SM1} + \varepsilon_{SM1}$	(3.11)
$y_D = C_D + \varepsilon_D$	(3.12)
$y_{SM2} = C_{SM2} + \varepsilon_{SM2}$	(3.13)
$y_P = C_P + \varepsilon_P$	(3.14)

Table 4. Initial conditions for center-point batch reactor operation at$T = 40$

State variable	Units	Initial condition
C_{SM1}	M	1
C_D	M	0.05
C_{SM2}	M	1.15
C_{SM1D}	M	0
C_P	M	0
C_{H_2O}	M	0.10
C_{I1}	M	0
C_{I2}	M	0
C_{I3}	M	0
C_{I4}	M	0

Table 5 shows the true kinetic coefficients used by Domagalski et al. for generating simulated data. These parameter values were used in the current study to compute true kinetic and equilibrium coefficients via Arrhenius expressions:

$$k_i(T) = k_{i,ref} \exp\left(-\frac{E_{a,i}}{R}\left(\frac{1}{T} - \frac{1}{T_{\text{ref}}}\right)\right) (17)$$

$$K_{\text{eq}}(T) = K_{eq,ref} \exp\left(-\frac{\Delta H_1}{R}\left(\frac{1}{T} - \frac{1}{T_{\text{ref}}}\right)\right) (18)$$

where k_i is the ith kinetic coefficient, R is the universal gas constant, T is the temperature in K, and$T_{\text{ref}} = 313.15$ K $= 40$ °C is a reference temperature. In equation (18), K_{eq} is the equilibrium coefficient for reaction (1), and ΔH_1 is the reaction enthalpy. Table 6 provides measurement noise variances used in this study for generating simulated data.[21] Figure 2 shows one set of simulated old data generated using the values in

Table **5** and Table 6. As shown in the simulated true response in Figure 2, consumption of catalyst D is initially very fast and then the catalyst gets released via reaction (2) as the product is formed.

Table 5. True values of the kinetic coefficients and equilibrium constant[6]

	Units	Value at 40 °C	$E_{a,i}$ or ΔH_1 $\left(\text{J.mol}^{-1}\right)$
k_{1f}	$M^{-1}s^{-1}$	1.09×10^{-01}	5.00×10^4
K_{eq}	M^{-1}	9.33	-1.00×10^4
k_2	$M^{-1}s^{-1}$	3.39×10^{-03}	4.25×10^4
k_3	$M^{-1}s^{-1}$	1.09×10^{-06}	9.00×10^4
k_4	$M^{-1}s^{-1}$	1.17×10^{-05}	9.50×10^4
k_5	$M^{-1}s^{-1}$	1.93×10^{-06}	9.75×10^4
k_6	s^{-1}	1.87×10^{-08}	8.00×10^4

Table 6. Measurement variances used for generating simulated data

Measured response (M)	$\sigma^2_{y_i}$ (M^2)
y_{SM1}	4.8×10^{-2}
y_D	2.1×10^{-4}
y_{SM2}	5.2×10^{-2}
y_P	5.0×10^{-2}

The initial case study assumes that the system operates at$T = T_{\text{ref}} = 40$, which results in seven parame-

ters requiring estimation (i.e., $\theta = [k_{1f},\ K_{\text{eq}},\ k_2,\ k_3,\ k_4, k_5,\ k_6]^T$). These seven parameters lead to seven columns in the scaled sensitivity matrix $\mathbf{Z}$. Decision variables for the new experiments are initial concentrations for the reactants $SM1$, D, and $SM2$(i.e.,$\mathbf{d} = [C_{\text{SM1}_0}, C_{D_0}, C_{\text{SM2}_0}]^T$). Lower and upper bounds for these decision variables are provided in Table 7. Note that, the four approaches (i.e., LO-LO, Bayes-LO, LO-Bayes and Bayes-Bayes) that are compared in this initial case study also considered in an expanded case study (reported in the Supplementary Information) where temperature is an additional decision variable ($35 \leq T \leq\ 45$). Considering T as an additional decision variable results in a model with 14 parameters.

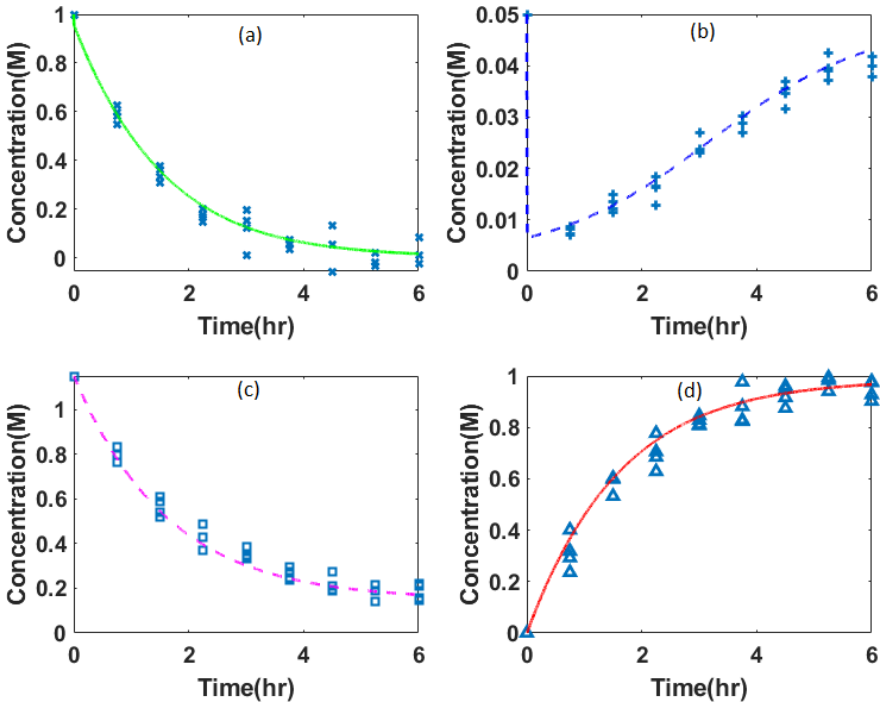

Figure 2. Simulated data (symbols) and noise-free concentration profiles (curves) obtained at 40 °C using experimental settings in Table 4, kinetic coefficients in

Table **5** and measurement errors in Table 6: a) SM1; b) D; c) SM2; d) P.

In the initial case study, three new sequential experiments are designed one-at-a-time. In the first step, one A-optimal experiment is designed and the parameters are estimated using both the old and the new data. Next, these parameter estimates and all of the data obtained are used to design a second sequential experiment. Finally, a third experiment is designed using parameter estimates and data from previous steps. For comparison, three new experiments are designed all-at-once based on the old data in Figure 2 and corresponding parameter estimates.

Table 7. Lower and upper bounds for the decision variables

Decision variables (M)	Lower bound (M)	Upper bound (M)
C_{SM1_0}	0.10	1.5
C_{D_0}	0.01	0.5
C_{SM2_0}	0.10	1.5

Table 8 provides information about the user-specified prior parameter information used in three different

Cases. The prior parameter guesses and corresponding standard deviations were used in the Bayesian objective function for parameter estimation (equation (6)). This prior information is also used in LO parameter estimation to obtain scaled sensitivity coefficients in $\mathbf{Z}$ (see equations (2) and (3)). As result, prior assumptions about parameters influence which parameters are estimated and which remained fixed at their initial values. The three cases described in Table 8 were used to investigate the influence of the prior parameter information on the quality of parameter estimates and experimental settings. For all three cases, parameter initial guess$\hat{\theta}_{j0}$ were selected randomly from normal distributions with true mean θ_j and true standard deviations_{θ_j} (see Table 5). In Case I, the modeler specifies prior information that is quite accurate (i.e., prior parameter standard deviations are 1/5 of the true value), whereas in Case II, the modeler is less certain about the initial parameter guesses. The selection rules in the third column of Table 8 for Cases I and II prevent random selection of unrealistic negative parameter values and parameter values more than 3 standard deviations from the true parameter values. Case III is used to investigate whether the Bayesian or LO approach to MBDOE and parameter estimation is more robust to misinformed prior information (i.e., when modelers mistakenly believe that they know more about the plausible parameter values than is warranted). In Case III, initial parameter guesses are further from the true values than the modeler believes.

Table 8. Selection of parameter initial guesses from normal distributions

	Prior Distribution for the jth parameter	**Rules for selection of p**
Case I: Informative initial guess	$N(\mu:\ \theta_j,\ \sigma:\ s_{\theta_j} = \frac{1}{5}\theta_j)$	Discard any parameter ini
Case II: Moderately informative initial guess	$N(\mu:\ \theta_j,\ \sigma:\ s_{\theta_j} = \frac{1}{2}\theta_j)$	Discard any negative para
Case III: Misinformed initial guess	$N(\mu:\ \theta_j,\ \sigma:\ s_{\theta_j} = \frac{1}{5}\theta_j)$	Select initial guess all **fro**

Monte Carlo Simulation Results and Discussion

Case I: Results when informative parameter initial guesses are used

In this Case, 100 initial guesses for the seven parameters were selected as described in the first row of Table 8. Using each set of initial guesses and the simulated old data in Figure 2, preliminary values of the model parameters were estimated using both Bayesian and LO approaches. All seven parameters were estimated using the Bayesian approach, whereas subsets of parameters were estimated using the LO approach, with remaining parameters fixed at their initial guesses. Using the LO approach, parameter k_2 was always ranked as the most-estimable parameter, followed by K_{eq}, k_{1f} andk_3 (using the ranking algorithm in Table 2). Parametersk_4, k_5 and k_6 were always left out of the ranked list. Using Wu's r_{cc} criterion, the parameter subset$\theta_{\text{sub}} = [k_2,\ K_{\text{eq}}]^T$was selected for estimation in all 100 simulated old data sets. Parametersk_{1f}, k_3, k_4, k_5 and k_6 were fixed at their initial values.

The preliminary parameter estimates obtained via Bayesian and LO estimation were then used to design sequential A-optimal experiments using Bayesian and LO approaches. Details concerning how many and which parameters tended to be estimated after each stage of sequential experimentation are provided in the Supplementary Information.

Figure 3 provides boxplots for 100 values of the scaled sum of squared deviations between the estimated and true parameter values:

$$\text{SS}D_\theta = \left(\ \hat{\theta} - \theta^{\text{true}}\right)^T \mathbf{\Sigma_0^{-1}} \left(\ \hat{\theta} - \theta^{\text{true}}\right) \quad (19)$$

for all four approaches when selecting three new A-optimal experiments, one at a time. The Bayes-LO approach is the superior approach on average, resulting in the smallest mean and median for$\text{SS}D_\theta$ after each round of experimentation. The results in Figure 3 indicate that designing experiments using the proposed

modified Bayesian approach (i.e., with equation (16) as the objective function) is superior to designing experiments using the LO approach (i.e., the Bayes-Bayes results are better than LO-Bayes, and the Bayes-LO results are better than LO-LO). In addition, parameter estimation using the LO approach is superior to the Bayesian approach (i.e., Bayes-LO is better than Bayes-Bayes, and LO-LO is better than LO-Bayes).

Figure 4 shows boxplots for 100 values of $\mathrm{SS}D_\theta$ when designing three new A-optimal experiments all at once for Case I. Bayes-LO is the superior approach and LO-Bayes is the worst approach. Comparing these results with the results in Figure 3, it can be concluded that designing experiments one-at-a-time resulted in better final parameter values than designing all three new experiments at once.

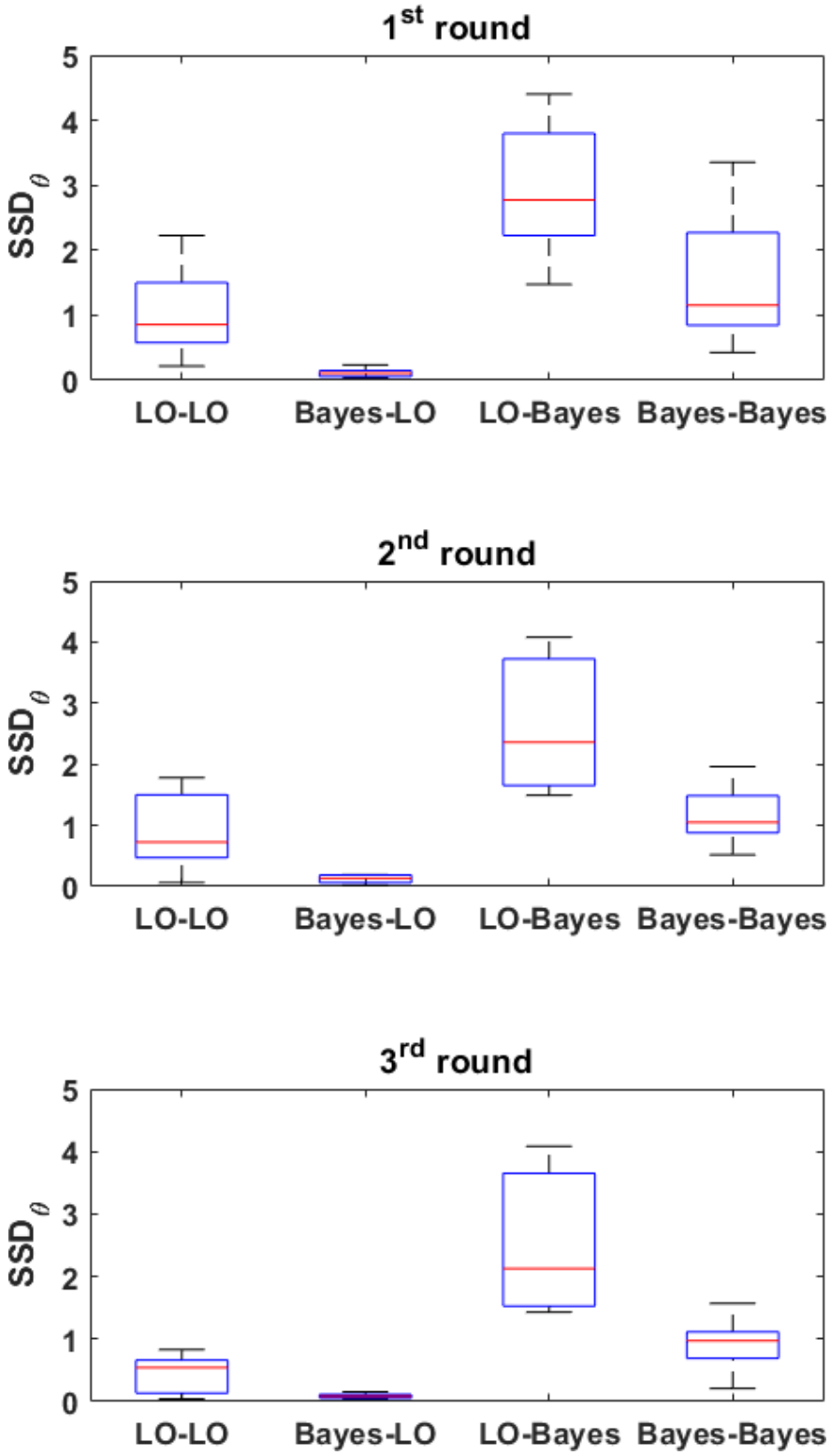

Figure 3 . Boxplots for 100 values of SSD_θ for Case I, when designing three sequential A-optimal experiments

one at a time using LO-LO, Bayes-LO, LO-Bayes, and Bayes-Bayes approaches

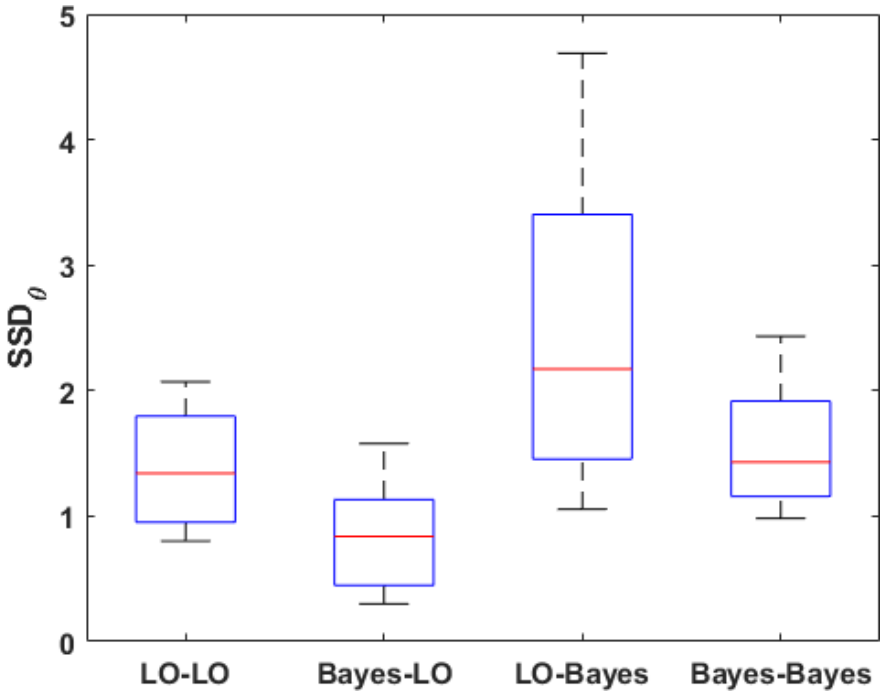

Figure 4. Boxplots for 100 values of SSD_θ for Case I, when designing three sequential A-optimal experiment all at once using Bayes-Bayes, LO-LO, Bayes-LO, and LO-Bayes approaches

Computation times for both Bayesian and LO approaches to MBDOE were compared using MC simulations for the Bayes-Bayes and LO-Bayes approaches. These two approaches use the same approach for parameter estimation, but different methods for MBDOE, making it possible to isolate the effects of Bayesian and LO approaches for MBDoE. Using a core i5 laptop with 8 GB RAM, the average computations time for each Bayes-Bayes run was 51.2 s, which is faster than 89.9 s on average for a LO-Bayes run. Although this difference is relatively small for the current case study, we anticipate that larger differences could occur for larger models with more parameters and decision variables.

Case II: Results when moderately-informative initial guesses are available

Figure 5 compares boxplots for the 100 values of SS$D_\theta$for four approaches when designing three A-optimal experiments one at a time using the prior parameter guesses in Case II. Similar patterns are observed compared to Case I: as more experiments are designed and more data become available, the mean and median values of boxplots forSSD_θ becomes smaller. However, since the parameter initial guesses are not as good as in Case I, the parameter estimates in Case II are less accurate than in Case I. As in Case I, the Bayes-LO approach provides the best parameter values on average. A key difference between the results in Case I and Case II is that the LO approach tended to estimate more parameters in Case II, due to higher initial parameter uncertainties. Details concerning the frequency with which different parameters were estimated are provided in the Supplementary Information.

Case III: Results when misinformed initial guesses are used

Figure 6 shows boxplots for 100 values of SS$D_\theta$obtained for Case III, with the misinformed parameter initial guesses described in the third row of Table 8. As expected, because parameter initial guesses were worse

than in Case I and II, the mean and median for SSD_θ are larger than Case I and II for all four approaches. The Bayes-LO approach to designing experiments and estimating parameters was the best approach, even though the modeler believed he or she had better prior knowledge about the parameters than was justifiable. These results suggest that the Bayes-LO approach is somewhat robust to specification of misinformed prior information. Note that the settings in Case II and Case III were also used to design three experiments all-at-once instead of one at a time. As in Case I, the Bayes-LO approach was the best and the parameter estimations resulting from the three-at-once experiments were not as good as those obtained using the one-at-a-time approach. Details are provided in the Supplementary Information.

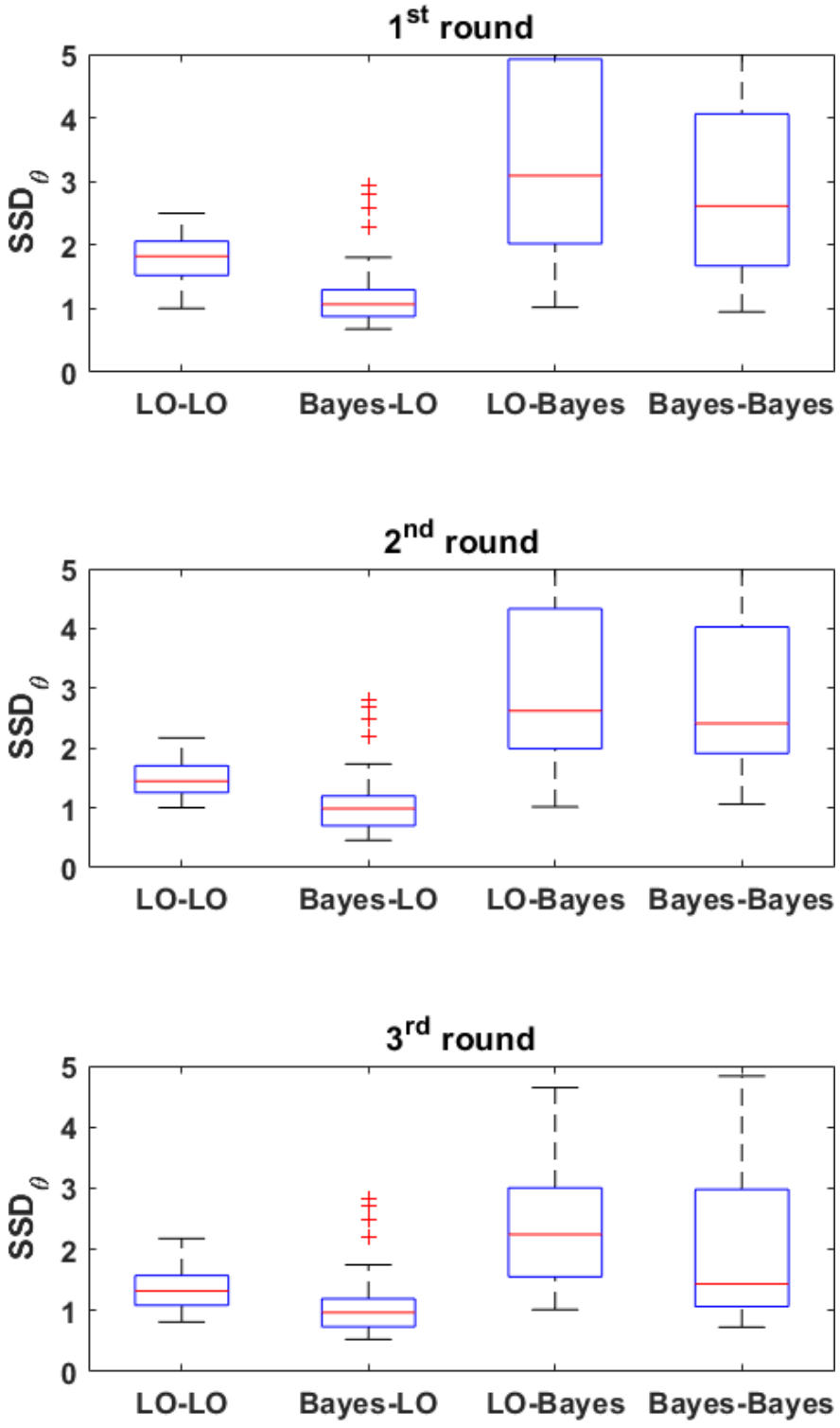

Figure 5. Boxplots for 100 values of SSD_θ for Case II, when designing three sequential A-optimal experiment one at a time using LO-LO, Bayes-LO, LO-Bayes, and Bayes-Bayes approaches

Cases I, II and III were also repeated using the extended 14 parameter model that arises (due to activation energies and reaction enthalpy) when the temperature is included as an additional decision variable. The same trends were observed as for the 7 parameters model, indicating that the Bayes-LO is the best of the four approaches studied. Details are provided in the Supplementary Information.

Conclusions

A simple Bayesian approach is proposed for the sequential model-based design of experiments (MBDoE) when the **FIM** is noninvertible. The results for the proposed Bayesian approach were compared with a leave-out (LO) approach developed in previous studies. In addition, the effectiveness of Bayesian and LO approaches for parameter estimation were also compared, so that four different approaches were investigated (i.e., Bayes-Bayes, LO-LO, Bayes-LO, and LO-Bayes) were investigated. These approaches were tested using simulated data generated from a 7-parameter isothermal pharmaceutical production model and a corresponding 14-parameter non-isothermal model. Three different cases were considered wherein the modeler specified different prior information about the parameters. The results indicate that the Bayes-LO approach (i.e., a Bayesian approach for MBDoE combined with a LO approach for parameter estimations) is superior to the three other approaches. The proposed Bayesian approach for designing experiments consistently provided superior experiments for use in parameter estimation compared with the LO approach. However, after new experimental data had been obtained, the LO approach for parameter estimation consistently provided parameter values that were closer, on average, to their true values than parameter estimates obtained from Bayesian estimation. Promising simulation results obtained using misspecified prior parameter knowledge indicated that the Bayes-LO approach was somewhat robust to misinformation for the current case study.

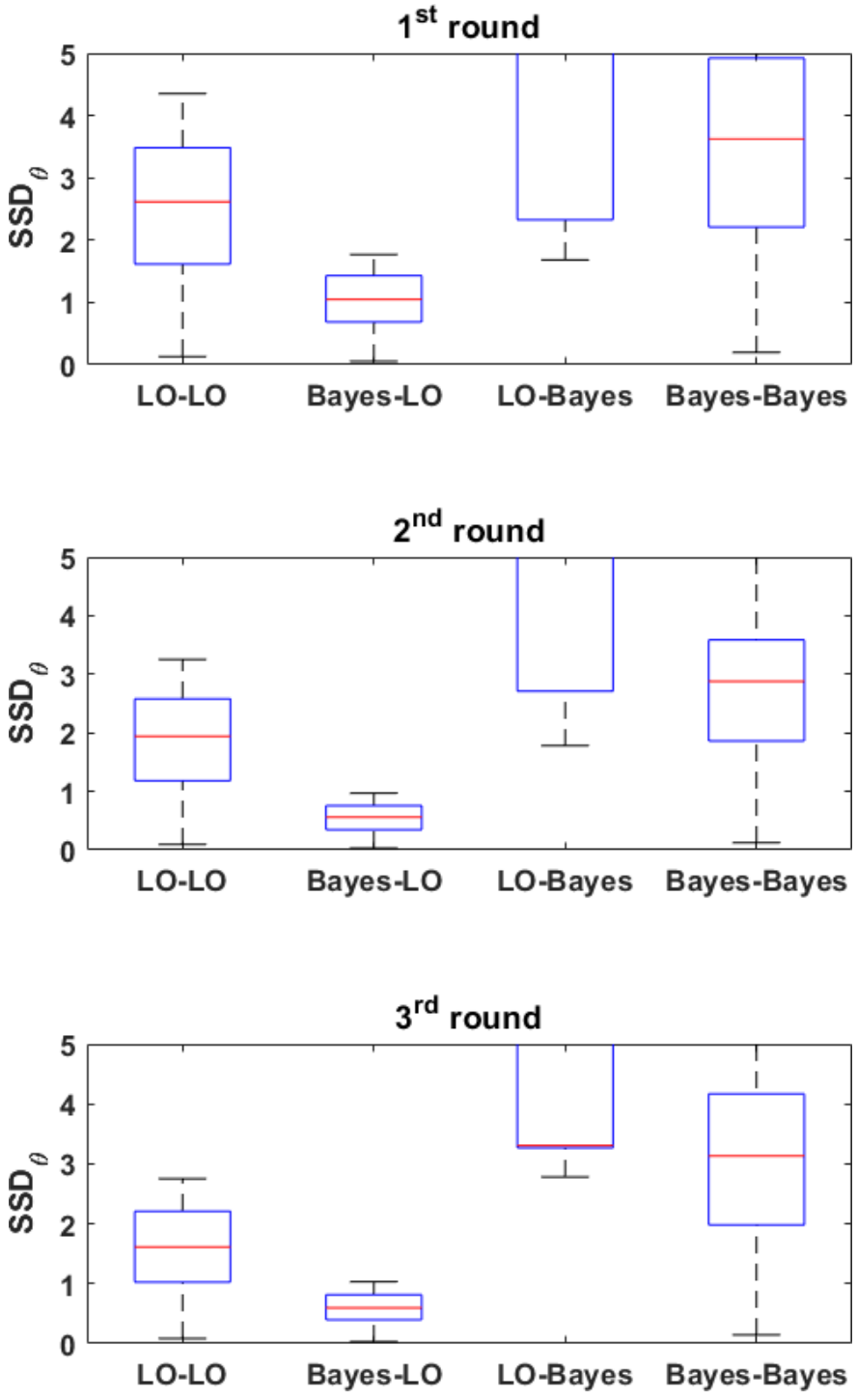

Figure 6. Boxplots for 100 values of SSD_θ for Case III, when designing three sequential A-optimal experiment one at a time using Bayes-Bayes, LO-LO, Bayes-LO, and LO-Bayes approaches

ACKNOWLEDGMENT

Authors acknowledge financial support from Eli Lilly and Company and technical advice from Dr. Salvador Garcia-Munoz and Dr. Brandon Jacob Reizman. An Ontario Trillium Scholarship for A. Shahmohammadi from government from the province of Ontario is also gratefully acknowledged.

References

1. Okino MS, Mavrovouniotis ML. Simplification of mathematical models of chemical reaction systems. *Chem Rev* . 1998. doi:10.1021/cr950223l

2. Dobre TG, Sanchez Marcano JG. *Chemical Engineering: Modelling, Simulation and Similitude* .; 2007. doi:10.1002/9783527611096

3. Birtwistle MR, Mager DE, Gallo JM. Mechanistic Vs. empirical network models of drug action. *CPT Pharmacometrics Syst Pharmacol* . 2013. doi:10.1038/psp.2013.51

4. Gernaey K V., Lantz AE, Tufvesson P, Woodley JM, Sin G. Application of mechanistic models to fermentation and biocatalysis for next-generation processes. *Trends Biotechnol* . 2010. doi:10.1016/j.tibtech.2010.03.006

5. Boukouvala F, Muzzio FJ, Ierapetritou MG. Dynamic data-driven modeling of pharmaceutical processes. *Ind Eng Chem Res* . 2011. doi:10.1021/ie102305a

6. Domagalski NR, Mack BC, Tabora JE. Analysis of Design of Experiments with Dynamic Responses. *Org Process Res Dev* . 2015. doi:10.1021/acs.oprd.5b00143

7. Feng AL, Boraey MA, Gwin MA, Finlay PR, Kuehl PJ, Vehring R. Mechanistic models facilitate efficient development of leucine containing microparticles for pulmonary drug delivery. *Int J Pharm* . 2011. doi:10.1016/j.ijpharm.2011.02.049

8. Kourti T, Lepore J, Liesum L, et al. Scientific and Regulatory Considerations for Implementing Mathematical Models in the Quality by Design (QbD) Framework. *Pharm Eng* . 2015.

9. Karlsson S, Rasmuson A, Van WaChem B, Björn IN. CFD modeling of the wurster bed coater. *AIChE J* . 2009. doi:10.1002/aic.11847

10. Maria G. A review of algorithms and trends in kinetic model identification for chemical and biochemical systems. *Chem Biochem Eng Q* . 2004;18(3):195-222.

11. Atkinson AC. Developments in the Design of Experiments, Correspondent Paper. *Int Stat Rev / Rev Int Stat* . 1982;50(2):161-177. doi:10.2307/1402599

12. Liu S, Neudecker H. A V-optimal design for Scheffé's polynomial model. *Stat Probab Lett* . 1995;23(3):253-258. doi:10.1016/0167-7152(94)00122-O

13. Chaloner K, Verdinelli I. Bayesian experimental design: A review.*Stat Sci* . 1995;10(3):273-304. doi:10.1214/ss/1177009939

14. Galvanin F, Barolo M, Bezzo F. Online model-based redesign of experiments for parameter estimation in dynamic systems. *Ind Eng Chem Res* . 2009;48(9):4415-4427. doi:10.1021/ie8018356

15. John RCS, Draper NR. D-Optimality for regression designs: A review. *Technometrics* . 1975;17(1):15-23. doi:10.1080/00401706.1975.10489266

16. Krafft O, Schaefer M. D-optimal designs for a multivariate regression model. *J Multivar Anal* . 1992;42(1):130-140. doi:10.1016/0047-259X(92)90083-R

17. Box GEP, Lucas HL. *Design of Experiments in Non-Linear Situations* . Vol 46. [Oxford University Press, Biometrika Trust]; 1959.

18. Wong WK. Comparing robust properties of A, D, E and G-optimal designs. *Comput Stat Data Anal* . 1994. doi:10.1016/0167-9473(94)90161-9

19. Stigler SM. Optimal experimental design for polynomial regression.*J Am Stat Assoc* . 1971;66(334):311-318.

20. Shahmohammadi A. Model-based optimal design of experiments with noninvertible fisher information matrix. 2019.

21. Shahmohammadi A, McAuley KB. Sequential Model-Based A- and V-Optimal Design of Experiments for Building Fundamental Models of Pharmaceutical Production Processes. *Comput Chem Eng* . 2019.

22. Franceschini G, Macchietto S. Model-based design of experiments for parameter precision: State of the art. *Chem Eng Sci* . 2008. doi:10.1016/j.ces.2007.11.034

23. López DC, Barz T, Peñuela M, Villegas A, Ochoa S, Wozny G. Model-based identifiable parameter determination applied to a simultaneous saccharification and fermentation process model for bio-ethanol production. *Biotechnol Prog* . 2013;29(4):1064-1082. doi:10.1002/btpr.1753

24. Walter E, Pronzato L. Qualitative and quantitative experiment design for phenomenological models-A survey. *Automatica* . 1990;26(2):195-213. doi:10.1016/0005-1098(90)90116-Y

25. Pinto JC, Lobão MW, Monteiro JL. Sequential experimental design for parameter estimation: a different approach. *Chem Eng Sci* . 1990;45(4):883-892. doi:http://dx.doi.org/10.1016/0009-2509(90)85010-B

26. Ford I, Silvey SD. A sequentially constructed design for estimating a nonlinear parametric function. *Biometrika* . 1980;67(2):381-388. doi:10.1093/biomet/67.2.381

27. Issanchou S, Cognet P, Cabassud M. Sequential experimental design strategy for rapid kinetic modeling of chemical synthesis. *AIChE J* . 2005;51(6):1773-1781. doi:10.1002/aic.10439

28. Bauer I, Bock HG, Körkel S, Schlöder JP. Numerical methods for optimum experimental design in DAE systems. *J Comput Appl Math* . 2000;120(1):1-25. doi:http://dx.doi.org/10.1016/S0377-0427(00)00300-9

29. López C. DC, Barz T, Körkel S, Wozny G. Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design.*Comput Chem Eng* . 2015;77:24-42. doi:10.1016/j.compchemeng.2015.03.002

30. Littlejohns J V, McAuley KB, Daugulis AJ. Model for a solid–liquid stirred tank two-phase partitioning bioscrubber for the treatment of BTEX. *J Hazard Mater* . 2010;175(1):872-882. doi:http://dx.doi.org/10.1016/j.jhazmat.2009.10.091

31. Ben-Zvi A, McAuley K, McLellan J. Identifiability study of a liquid-liquid phase-transfer catalyzed reaction system. *AIChE J* . 2004;50(10):2493-2501. doi:10.1002/aic.10202

32. Zhao YR, Arriola DJ, Puskas JE, McAuley KB. Applying multidimensional method of moments for modeling and estimating parameters for arborescent polyisobutylene production in batch reactor.*Macromol Theory Simulations* . 2017;26(1). doi:10.1002/mats.201600004

33. Issanchou S, Cognet P, Cabassud M. Precise parameter estimation for chemical batch reactions in heterogeneous medium. *Chem Eng Sci* . 2003;58(9):1805-1813. doi:10.1016/S0009-2509(03)00004-6

34. Cho KH, Shin SY, Kolch W, Wolkenhauer O. Experimental Design in Systems Biology, Based on Parameter Sensitivity Analysis Using a Monte Carlo Method: A Case Study for the TNF -Mediated NF- B Signal Transduction Pathway. *Simulation* . 2003. doi:10.1177/0037549703040943

35. Thompson DE, McAuley KB, McLellan PJ. Design of optimal sequential experiments to improve model predictions from a polyethylene molecular weight distribution model. *Macromol React Eng* . 2010;4(1):73-85. doi:10.1002/mren.200900033

36. Greville TNE. The pseudoinverse of a rectangular or singular matrix and its application to the solution of systems of linear equations.*SIAM Rev* . 1959;1(1):38-43. doi:10.1137/1001003

37. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. *J R Stat Soc Ser B (Statistical Methodol* . 2011;73(3):273-282. doi:10.1111/j.1467-9868.2011.00771.x

38. Johansen TA. On Tikhonov regularization, bias and variance in nonlinear system identification. *Automatica* . 1997;33(3):441-446. doi:http://dx.doi.org/10.1016/S0005-1098(96)00168-9

39. Tikhonov AN, Goncharsky A V., Stepanov V V., Yagola AG. Numerical methods for the solution of ill-posed problems. *Math Comput* . 1978;32(144):1320. doi:10.2307/2006360

40. Hoerl AE, Kennard RW. Ridge Regression: Biased estimation for nonorthogonal problems. *Technometrics* . 1970;12(1):55. doi:10.2307/1267351

41. Durán MA, White BS. Bayesian estimation applied to effective heat transfer coefficients in a packed bed. *Chem Eng Sci* . 1995;50(3):495-510. doi:http://dx.doi.org/10.1016/0009-2509(94)00260-X

42. Ruggoo A, Vandebroek M. Bayesian sequential script dsign optimal model-robust designs. *Comput Stat Data Anal* . 2004;47(4):655-673. doi:10.1016/j.csda.2003.09.014

43. Kravaris C, Hahn J, Chu Y. Advances and selected recent developments in state and parameter estimation. *Comput Chem Eng* . 2013;51:111-123. doi:10.1016/j.compchemeng.2012.06.001

44. Barz T, López Cárdenas DC, Arellano-Garcia H, Wozny G. Experimental evaluation of an approach to online redesign of experiments for parameter determination. *AIChE J* . 2013. doi:10.1002/aic.13957

45. Yao KZ, Shaw BM, Kou B, McAuley KB, Bacon DW. Modeling Ethylene/Butene copolymerization with multi-site catalysts: parameter estimability and experimental design. *Polym React Eng* . 2003;11(3):563-588. doi:10.1081/PRE-120024426

46. Shahmohammadi A, McAuley KB. Sequential Model-Based A-Optimal Design of Experiments When the Fisher Information Matrix Is Noninvertible.*Ind Eng Chem Res* . 2019;58(3):1244-1261. doi:10.1021/acs.iecr.8b03047

47. Dube MA, Penlidis A, Reilly PM. A systematic approach to the study of multicomponent polymerization kinetics: The butyl acrylate/methyl methacrylate/vinyl acetate example. IV. Optimal Bayesian design of emulsion terpolymerization experiments in a pilot plant reactor. *J Polym Sci Part A Polym Chem* . 1996. doi:10.1002/(SICI)1099-0518(19960415)34:5<811::AID-POLA11>3.3.CO;2-3

48. Hsu SH, Stamatis SD, Caruthers JM, et al. Bayesian framework for building kinetic models of catalytic systems. *Ind Eng Chem Res* . 2009. doi:10.1021/ie801651y

49. Han C, Chaloner K. Bayesian Experimental Design for Nonlinear Mixed-Effects Models with Application to HIV Dynamics.*Biometrics* . 2004. doi:10.1111/j.0006-341X.2004.00148.x

50. Van Den Berg J, Curtis A, Trampert J. Optimal nonlinear Bayesian experimental design: An application to amplitude versus offset experiments. *Geophys J Int* . 2003. doi:10.1046/j.1365-246X.2003.02048.x

51. Jaakkola TS, Jordan MI. Bayesian parameter estimation via variational methods. *Stat Comput* . 2000. doi:10.1023/A:1008932416310

52. Lehninger. *Principles of Biochemistry, 6th Ed.* ; 2009. doi:10.1017/CBO9781107415324.004

53. Petersen B, Gernaey K, Vanrolleghem PA. Practical identifiability of model parameters by combined respirometric-titrimetric measurements. In: *Water Science and Technology* . ; 2001.

54. Thompson DE, McAuley KB, McLellan PJ. Parameter estimation in a simplified MWD model for HDPE produced by a ziegler-natta catalyst.*Macromol React Eng* . 2009;3(4):160-177. doi:10.1002/mren.200800052

55. Lin W, Biegler LT, Jacobson AM. Modeling and optimization of a seeded suspension polymerization process. *Chem Eng Sci* . 2010;65(15):4350-4362. doi:10.1016/j.ces.2010.03.052

56. Wu S, McLean K a. P, Harris TJ, McAuley KB. Selection of optimal parameter set using estimability analysis and MSE-based model-selection criterion. *Int J Adv Mechatron Syst* . 2011;3(3):188. doi:10.1504/IJAMECHS.2011.042615

57. Mclean KAP, Mcauley KB. Mathematical modelling of chemical processes-obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. *Can J Chem Eng* . 2012;90(2):351-366. doi:10.1002/cjce.20660

58. Wu S, Mcauley KB, Harris TJ. Selection of simplified models: II. Development of a model selection criterion based on mean squared error.*Can J Chem Eng* . 2011;89(2):325-336. doi:10.1002/cjce.20479

59. Chu Y, Hahn J. Parameter set selection via clustering of parameters into pairwise indistinguishable groups of parameters. *Ind Eng Chem Res* . 2009. doi:10.1021/ie800432s

60. LORD FM. MAXIMUM LIKELIHOOD AND BAYESIAN PARAMETER ESTIMATION IN ITEM RESPONSE THEORY. *J Educ Meas* . 1986. doi:10.1111/j.1745-3984.1986.tb00241.x

61. Bates DM, Watts DG. Review of Linear Regression. *Nonlinear Regres Anal Its Appl* . 1988:1-31. doi:10.1002/9780470316757.ch1

62. Fabian B, Bernd H. Uncertainty estimation for linearised inverse problems comparing Bayesian inference and a pseudoinverse approach for acoustic transmission measurements. *tm - Tech Mess* . 2017;84:217. doi:10.1515/teme-2016-0022

63. Englezos P, Kalogerakis N. *Applied Parameter Estimation for Chemical Engineers* . Vol 53.; 1989. doi:10.1017/CBO9781107415324.004

64. Bard Y. Nonlinear Parameter Estimation. *Oper Res Q 19701977* . 1974;49(3):341.

65. Ryan KJ. Estimating Expected Information Gains for Experimental Designs with Application to the Random Fatigue-Limit Model. *J Comput Graph Stat* . 2003. doi:10.1198/1061860032012

66. Cox DR, Reid N. *The Theory of the Design of Experiments* . Chapman and Hall/CRC; 2000.

67. DuMouchel W, Jones B. A simple bayesian modification of d-optimal designs to reduce dependence on an assumed model. *Technometrics* . 1994. doi:10.1080/00401706.1994.10485399

68. Vivaldo-Lima E, Penlidis A, Wood PE, Hamielec AE. Determination of the relative importance of process factors on particle size distribution in suspension polymerization using a bayesian experimental design technique. *J Appl Polym Sci* . 2006. doi:10.1002/app.24889

69. Lindley D V. On a Measure of the Information Provided by an Experiment. *Ann Math Stat* . 1956;27(4):986-1005. doi:10.1214/aoms/1177728069

70. Papadimitriou C, Argyris C. Bayesian optimal experimental design for parameter estimation and response predictions in complex dynamical systems. In: *Procedia Engineering* . ; 2017. doi:10.1016/j.proeng.2017.09.205

71. Alexanderian A, Gloor PJ, Ghattas O. On Bayesian A- and D-optimal experimental designs in infinite dimensions. *Bayesian Anal* . 2016. doi:10.1214/15-BA969

72. Papadimitriou C, Beck JL, Au SK. Entropy-based optimal sensor location for structural model updating. *JVC/Journal Vib Control* . 2000. doi:10.1177/107754630000600508

73. Sacks J, Welch WJ, Mitchell TJ, Wynn HP. Design and Analysis of Computer Experiments. *Stat Sci* . 1989;4(4):409-423.