A New Approach to Batch Process Optimization Using Experimental Design

Paul J. Wissmann and Martha A. Grover
School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA 30332

DOI 10.1002/aic.11715
Published online December 4, 2008 in Wiley InterScience (www.interscience.wiley.com).

Empirical and mechanistic experimental design methods are combined to construct partial models, which are, thus, used to design a process. The grid algorithm restricts the next experimental point to potential process optima, according to the confidence intervals around the optimal points, and works with any experimental design algorithm such as D-optimal. Two case studies show the advantages of implementing the grid algorithm. On average the improvement due to the grid algorithm was 15–20% in the first case study. The second case study is based on thin film growth using four potential models, with the most probable model used for experimental design. The grid algorithm balances the trade-off between two extremes: D-optimal designs and sampling at the predicted optimal point. The methodology presented shows that the experimenter does not have to decide ahead of time on purely empirical or mechanistic experimental design methods, since both may be useful. © 2008 American Institute of Chemical Engineers
AIChE J, 55: 342–353, 2009
Keywords: experimental design, modified Himmelblau function, partial models, prediction variance, grid algorithm, nucleation density, process design, batch process

Introduction

When finding the best settings for a batch process to create the desired amount and quality of product, one rarely has a perfect model of the process to ensure the correct settings are chosen. As a result, one performs experiments to relate process settings to final performance and properties. The experimental data is then combined with a model to estimate unknown model parameters. This model is then used to select additional experiments, and the model can be used to design improvements to the current process. In many cases of experimental design, an empirical model such as a polynomial fit is used.[1] In other cases, mechanistic models based on physical principles are used.[2] An example of a mechanistic model is an Arrhenius equation, which uses activation energies to model the temperature dependence of a reaction rate. A good mechanistic model can help an experimenter interpolate more accurately or even extrapolate from the available data. An empirical model, on the other hand, is made to fit the available data as efficiently as possible, without necessarily having any knowledge of the underlying phenomena. Because of an empirical model's best-fit nature, in general it cannot be used to predict the outcome of an experiment done outside of the sample data used to fit the model. Empirical and mechanistic models are not always developed separately, such as with semimechanistic[3] and grey-box models.[4] These two studies are focused on developing models from existing data, while very little effort has been given to experimental design to improve the models.

In process design and optimization, a systems engineer will often use a model that has already been developed by another researcher or engineer. Unfortunately, such models for new processes take years to develop. Rather than wait until a highly accurate mechanistic model has been developed, a systems engineer can use partial mechanistic models. A partial model will have limited accuracy and may still be under development, but it may still be useful in developing a new process, especially when restricted to a limited range of

Correspondence concerning this article should be addressed to M. Grover at martha.grover@chbe.gatech.edu.

process settings. With the ever-changing product industry, such as microelectronics and pharmaceuticals, one will have to modify and reoptimize a process many times over the equipment's lifetime. Not only could the process be used for different products, but the specifications for products could change frequently due to new technology or tighter government regulation. The advantage of a mechanistic model compared to an empirical fit is that it can be adapted and reused for such modified situations. Due to better understanding of the process, the time needed to perform subsequent optimizations can be reduced. However, empirical models are typically simpler to generate, and have fewer parameters to identify. If limited experimental data is available, they may be more useful for process optimization. We are proposing methods that use both types of models. One method is a special case of G-optimal design called P-optimal design, which is designed to improve the prediction at the optimal point. The second method called the grid algorithm reduces the design space using the confidence interval of the most probable model to concentrate experiments around the predicted optimal point.

There have been many articles written on the topics of model discrimination and parameter identification. A method for model discrimination developed by Box and coworkers[5] uses Bayesian probability to predict which mechanistic model is more probable, given existing experimental data. Equation 1 can be used for either mechanistic or empirical models

$$P(M_j|Y, MSE_j) = P(M_j) \times 2^{-p_j/2} \times MSE_j^{-v_e/2} \quad (1)$$

where M_j is model number j, Y is the experimental data, v_e is the number of repetitions for each data point, p_j is the number of parameters in model j, and MSE_j is the mean-squared error

$$MSE = \frac{\sum_{i=1}^{n} \left(y(x_i) - \hat{y}_j(x_i)\right)^2}{n} \quad (2)$$

where $y(x_i)$ is the experimental measurement, $i = 1, 2, \ldots, n$, and $\hat{y}_j(x_i)$ is the prediction from model j at point i. A model is penalized in Eq. 1 for having more parameters, fewer repetitions of the experiment, or larger error than an alternative model. This method is used to discriminate between existing models using existing data, but is not designed as an iterative process for experimental design. A sequential experimental scheme was developed to discriminate between multiple models[6] by designing the experiments where the models in question have the largest difference in prediction: $max_x(abs(\hat{y}_1(x) - \hat{y}_2(x)))$. Points that maximize the models' prediction differences, however, are not necessarily desirable when one is trying to gain information about the optimal point of the process. For example, the models might actually agree at the optimal point.

Work that combines experimental design with mechanistic and empirical models has been largely limited to studies for speeding up simulation times. Specifically, the concept of surrogate models has been introduced to replace complex mechanistic models with simpler empirical models.[7,8] Nonsimulation-based research in experimental design is largely separated depending on whether the model used is empirical or mechanistic. Extensive work in experimental design has been done for estimating the parameters in a mechanistic model using the D-optimal criterion to direct experiments with the objective of reducing the uncertainty in the parameters of a single model.[2,9] Specifically, D-optimal design does this by maximizing $|A'A|$, where A is the design matrix of a model. D_s-optimal is a modification to D-optimal, in which the determinant of the covariance matrix of a subset of the parameters is minimized.[10] Work by Franceschini et al. used experimental design to elucidate the parameters of kinetic models for a biodiesel process.[11] The drawback for D-optimal designs is the ability to only consider one model. If one is characterizing a new process, an established model may not be available.

In contrast, traditional design of experiments using factorial or fractional factorial designs is commonly used for optimizing a process based on empirical models.[12] An example of empirical modeling is response surface modeling, used in factorial design of experiments.[13] In this method, the objective is to maximize performance by finding the best operating point. Models developed using the Taguchi method are generally empirical as well.[14] The Taguchi method is an experimental design aimed at improving the quality of the product. Little effort is given to understanding the physics behind the results when these empirical approaches are used, but empirical models can be important and useful for finding the optimal operating point for a process quickly and efficiently.

Other techniques have been used to create a better sampling scheme. Defining a regular grid on the experimental space and randomly picking points from that grid is called Latin Hypercube sampling (LHS).[15] Alternatively, one can space the grid points irregularly based on spatial variation of the function, or adaptively based on previous samples and an experimental design objective.[16] All of these methods are designed for better sampling of the entire experimental region, whereas here we are interested in designing our experiments for best prediction at the unknown optimal point of a process.

Methodology

The focus here is to combine the best features of various methods for experimental design that are currently being used. By combining the methods for empirical and mechanistic models, one could efficiently find an optimal point, while also building mechanistic models that are useful in explaining the phenomena behind the process. Here optimal is defined as minimizing or maximizing some predefined objective function, but other definitions of optimal are also possible, such as designing the process to perform within a certain interval. The insight gained from these experiments can be used for future work, whether it is to reoptimize the process around a different operating regime, or to design a new process. At best, one or more of the models will have a good prediction over the entire region, such that it can be used for process design over the entire experimental range in question, although this will not necessarily be achieved.

The basic steps in our proposed methodology are illustrated by Figure 1. One begins with an initial probability of each model M_j, where $j = 1, 2, \ldots, m$, and a hypothesis as to

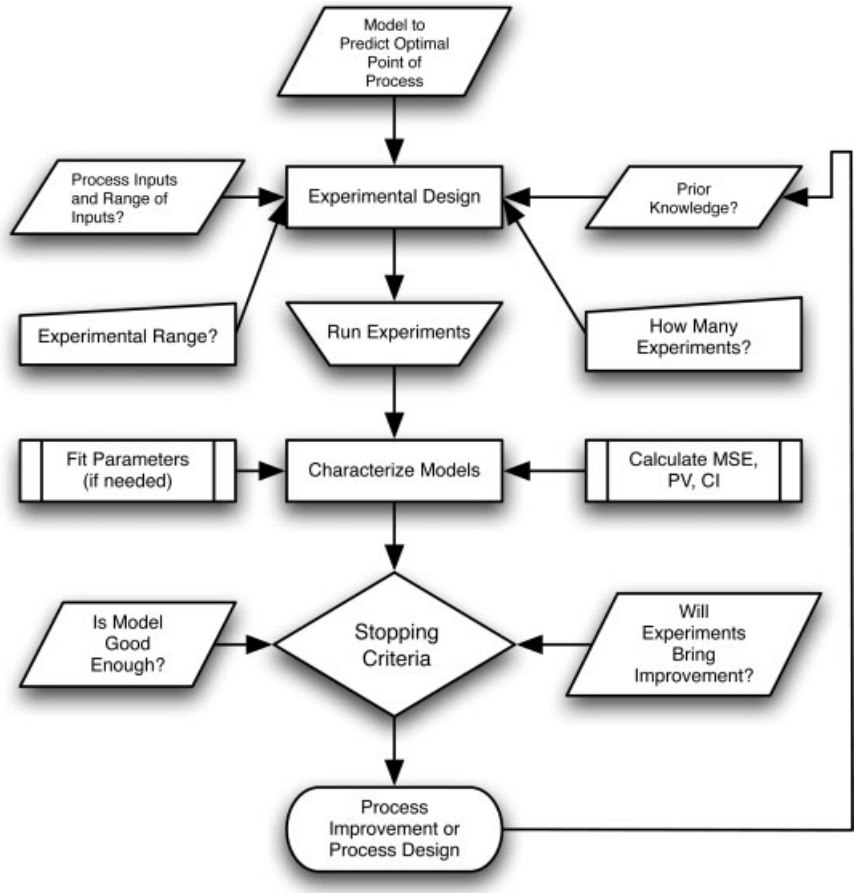

Figure 1. Proposed experimental design algorithm.

which process settings (temperature, pressure, etc.) should be varied to reach the design goal. The next step is to design an experiment or set of experiments to acquire some initial information about the system to obtain initial estimates of the unknown parameters θ_j in each model.

After this first set of experiments has been performed, we evaluate several types of experimental designs for the sequential portion. The three methods compared in this work are D-optimal $(D = \max_x |A_j' A_j|)$,[12] a random design of experiments, and our newly proposed P-optimal design which utilizes the prediction variance. A_j is defined as

$$A_j = \begin{pmatrix} a_{11}^{(j)} & a_{21}^{(j)} & \cdots & a_{pj1}^{(j)} \\ a_{12}^{(j)} & a_{22}^{(j)} & \cdots & a_{pj2}^{(j)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n}^{(j)} & a_{2n}^{(j)} & \cdots & a_{pjn}^{(j)} \end{pmatrix} \tag{3}$$

The prediction variance $\sigma_j^2(x)$ of model j reflects the uncertainty of the model prediction $\hat{y}_j(x)$, which is assumed here to be a scalar[17]

$$\sigma_j^2(x) = a^{(j)} (A_j' A_j)^{-1} a^{(i)\prime} \hat{\sigma}_j^2 \tag{4}$$

where

$$a_{lr}^{(j)} = [\partial \hat{y}_j(\theta_j, \mathbf{x}_r)/\partial \theta_{jl}]_{\theta_j = \hat{\theta}_j} \quad l = 1, \ldots, p_j \tag{5}$$

and $a_r^{(i)} = [a_{1r}, a_{2r}, ..., a_{p_j r}]$ are for the rth experimental setting, and $\theta_j = [\theta_{j1}, \theta_{j2}, ..., \theta_{jp_j}]$ is a vector of parameters for the j^{th} model. The model variance $\hat{\sigma}_j^2$ used in Eq. 4 is calculated similarly to MSE_j in Eq. 2 except the summation is divided by $n-p_j$, where p_j is the number of estimated parameters in model j, because p_j degrees of freedom are lost by estimating p_j parameters

$$\hat{\sigma}_j^2 = \frac{\sum_{i=1}^{n} (y(x_i) - \hat{y}_j(x_i))^2}{n - p_j} \tag{6}$$

D-optimal and P-optimal are experimental designs with different objectives. D-optimal designs are frequently used for experimental design and minimize the generalized variance of the parameter estimates. In contrast, G-optimal designs attempt to minimize the maximum variance of the predicted values y over the entire experimental region,[12] but are generally very computationally expensive and are less commonly used. Our P-optimal design is a modification of G-optimal. Rather than minimizing $\sigma_j^2(x)$ over the entire design region, P-optimal aims to minimize $\sigma_j^2(x)$ at a single point x^* $(\mathrm{P} = \min_{x_e} \sigma_j^2(x^*))$, where x_e is the experimental

point. Thus, P-optimal picks the experimental point which should most reduce $\sigma^2_{j^*}(x^*)$. In this work, $x^* = \hat{x}_{j^*}$, the predicted optimal using the most probable model j^*, based on the Bayesian probability of Eq. 1. Our rationale is that by focusing on $\hat{x}_{j^*}$, one can obtain a better prediction around that point. Specifically, the P-optimal algorithm tries different experimental points by adding an additional row to A_{j^*}, and finds the experimental point where $\sigma^2_{j^*}(\hat{x}_{j^*})$ is minimized. By performing an experiment at the point that minimizes $\sigma^2_{j^*}(\hat{x}_{j^*})$, the new experimental data will add confidence to the prediction $\hat{y}_{j^*}(\hat{x}_{j^*})$. One could use all j models, and weight their predictions by P_j to compute the expected value of $\hat{x}$ and $\hat{y}(\hat{x})$, but if there are multiple local minimum, the resulting intermediate points might lead to very poor designs. Therefore, in this study the most probable model is always used to design the next experiment.

After each experiment has been performed, the parameters for each model are re-estimated using a parameter estimation technique. In this work, the parameters are estimated using a least-squares minimization between the model and all of the data. Once the parameter estimate $\hat{\theta}$, has been obtained for each model, the performance of each model can then be examined. The optimal point for each model is estimated as

$$\hat{x}_j = \min_x f(\hat{y}_j(x)) \tag{7}$$

where $f(x)$ is the objective function to be minimized. Note that $\hat{x}_j$, is the estimated optimal point from the *model j*, and not necessarily $\bar{x}$, the true optimal point of the *process*. As a model becomes more accurate $\hat{x}_j$, and $\bar{x}$, will ideally converge to the same value.

The probability of each model is calculated using Eq. 1, and here we set the *a priori* probabilities $P(M_j)$ to be equal for all models. The model variance $\sigma^2_j(x)$ is used to calculate the confidence interval on $\hat{y}_j(x)$

$$CI_j(x) = \pm t_{\alpha/2,n-p_j}\sqrt{\sigma^2_j(x)} \tag{8}$$

where α is the level of confidence desired.[12] Once the performance of each model has been calculated, the stopping criteria are used to evaluate whether to continue the experiments or not. The stopping criteria used here are evaluated using the most probable model j^*. The first criterion considers whether the change in *MSE* of model j^* is less than ε_{tol}, where ε_{tol} is a prespecified constant value. The second criterion checks whether the confidence interval at $\hat{x}_{j^*}$ is below the desired level, ε_{CI}. If the first stopping criterion is true, then additional experiments are unlikely to improve the design or the confidence interval on the model prediction. If either criterion is true, then the experiments are finished and intervention by the experimenter is needed to interpret the results. The experimenter must decide whether the most probable model is good enough, if this model needs modification, or if an entirely new model is needed. At any time during the experiments, it is possible to add one or more new models and continue to iterate through the methodology. As part of our methodology development we also consider another new feature, which we call the grid algorithm. Its purpose is to further concentrate the experiments near $\hat{x}_{j^*}$ using the confidence interval of the objective function

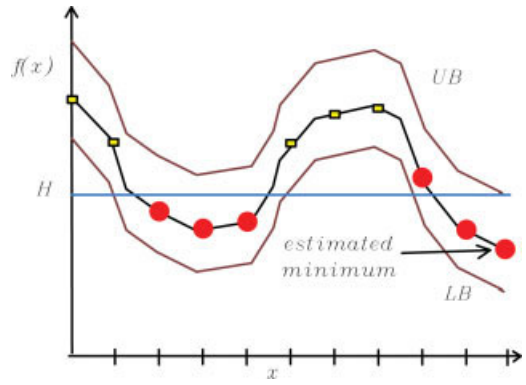

Figure 2. Graphical representation of the grid algorithm.

Filled circles are potential minima, while open squares are not. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

$CI(f(\hat{y}_{j^*}(\hat{x}_{j^*})))$. This is critical when partial models are being used, and is not necessarily achieved by D-, G-, or P-optimal, depending on the model or models being considered. The grid algorithm can be divided into five steps:

1. Defining a set of grid points x_g, in the experimental domain Z.
2. Calculating a threshold value $H(\hat{x}_{j^*})$

$$H(\hat{x}_{j^*}) = f(\hat{y}_{j^*}(\hat{x}_{j^*})) + CI(f(\hat{y}_{j^*}(\hat{x}_{j^*}))) \tag{9}$$

3. Restrict consideration to grid points with outputs that may be below the threshold value $f[\hat{y}_{j^*}(x_g)] - CI(f(\hat{y}_{j^*}(\hat{x}_g))) < H(\hat{x}_{j^*})$ (These are the potential optimal points, at the α confidence level.)
4. Calculate the output of the experimental design function $f_{ED}(x)$, at these remaining grid points and pick the best point $x_{e,o}$, from this set.
5. Find the optimal experimental point x_e, using $x_{e,o}$ as a starting point, constraining the search within a box centered at $x_{e,o}$, with sides equal to twice the grid spacing.

Equation 9 represents the case where the objective is to find the minimum value of $f(x)$. The lower bound of a prediction at point x is: $LB(x) = f(\hat{y}_{j^*}(x) - CI(f(\hat{y}_{j^*}(x)))$, and the upper bound is: $UB(x) = f(\hat{y}_{j^*}(x) - CI(f(\hat{y}_{j^*}(x)))$. In this case the confidence interval is added to the optimal value because one does not want to eliminate values of x in which $LB(x) < UB(\hat{x}_{j^*})$. This is shown graphically in Figure 2, in which all values of x having $LB < H$ are potentially the minimum (marked by circles). The points marked by squares are not potential minimum at a confidence level of α. Similar calculations can be performed when the objective is to find the maximum value.

In this work, only two experimental parameters are explored, making the grid method easy to implement. For higher dimensions of the experimental parameters, a full gridding of the design space may not be practical. In such cases, a stochastic sampling method such as simulated annealing[18] or multistart[19] could alternatively be used to obtain a sampling of the system over the experimental space in the regions of interest. To get parameters that fit better in

the region of interest, one may also use the grid algorithm to limit parameter fitting only to points that are below $H(\hat{x}_{j^*})$.

Description of Case Studies

In this work, two case studies are performed. The first case study is used to test the method on a simple system that is linear in the parameters. The second case study is used to test the method on a system of technological interest and uses several candidate models. Three models are purely empirical, and one model is mechanistic. This case study was chosen to reflect the ongoing experiments in our research group dealing with a chemical vapor deposition process.[20]

The grid algorithm was implemented using D-optimal, P-optimal, and a random selection of experiments. First, the D-optimal method is used to sequentially design the experiments, and parameter fits are made based on these experiments. At each iteration, the most probable model is used to predict the next experimental point. The entire methodology simulation is performed one hundred times at each noise level, and the results are averaged. This process is then repeated using P-optimal, and then using a random selection of experiments. In the case of random selection of experiments, the grid algorithm first creates the list of possible optimal points. A point from this list is then selected randomly as the next experiment.

MHF case study

We first evaluate the method using the modified Himmelblau function (MHF)[21]

$$y^{MHF}(x) = x_1^4 + x_2^4 - 21x_1^2 + 2x_1^2x_2 + 2x_1x_2^2 - 13x_2^2 - 13x_1 - 19x_2 + 227 \tag{10}$$

We consider a range of $[-5,5]$ for both x_1 and x_2. In this case study $f(x)$ is the same as $y(x)$. The minimum of Eq. 10 is at the point $x_1 = -3.80$, and $x_2 = -3.32$, with a corresponding function value of 43.3. The other local minima have function values of 63.5, 54.9, and 65.9. A contour plot of this function is shown in Figure 3.

This function is chosen because it possesses desirable properties for evaluating the experimental design method, including having a minimum in the interior and having multiple local minima. Equation 10 is used here to generate the simulated data, while a slightly different model is used to fit the sampled data

$$\hat{y}_1(x) = x_1^4 + x_2^4 - 21x_1^2 + 2x_1^2x_2 + \theta_1x_1x_2^2 - 13x_2^2 + \theta_2x_2 + \theta_3 \tag{11}$$

Equation 11 is missing the linear x_1 term from the original MHF function to create model mismatch, such that no values of the three fitted parameters (θ_{1-3}) will completely match all points. No model is ever perfect, so in practice there will always be some mismatch. Experimental design methods must be robust to these modeling errors. The whole methodology is also run with no model mismatch to ensure convergence with the various noise levels. In this case, all experimental designs converged to the true value and location of the optimal point.

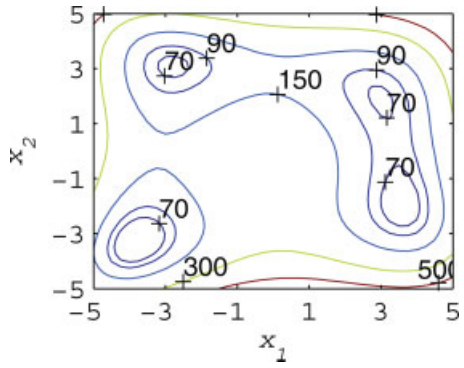

Figure 3. Modified Himmelblau function over $x_1 \in [-5, 5]$ and $x_2 \in [-5, 5]$.

[Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

The initial experiments for this case study are chosen to be a 2^2 factorial design (one experiment performed at each corner of the design space), plus an experiment at the center of the design space with two repetitions, for a total of six experiments ($2^2 + 2 \times (0,0)$. The center point is needed due to the nonlinear nature of the system being modeled, since there are quadratic terms in the MHF. The noise added (γ) to $y^{MHF}(x)$ is Gaussian and zero-mean, with standard deviations of 3, 15, or 30. The desired confidence interval (ε_{CI}) on the prediction is set to 5. The initial experiments are run, $\hat{\theta}_j$ is calculated, and the initial Bayesian probability is calculated using Eq. 1, as done in existing methods.

The MHF case study is used to examine three of the main choices affecting the D- and P-optimal and the grid method designs. First, the effect of the initial experiments is compared using our original experimental design ($2^2 + 2 \times (0,0)$, and the experimental design from the Latin Hypercube.[15] Second, the size of the grid for the grid algorithm at three different levels is compared. Third, the effect of the tolerance on the stopping criteria is investigated.

Film growth case study

This case study is based on a chemical vapor deposition process. In CVD, a thin film is grown on a heated substrate by flowing an evaporated metalorganic precursor over the substrate.[22] CVD experiments and film characterization can be time-consuming and costly, making CVD a good candidate for experimental design. Thin films grow after an initial nucleation of clusters on the substrate surface, and the density of nucleation sites on the substrate influences the film morphology and properties.[23] The purpose of this case study is to evaluate our experimental design method using multiple partial models. In this case study, the method picks the best of four possible models that predict the flux of atoms to the substrate surface, given the *simulated* experimental data. The two experimental settings considered are $x = [T,C]$. T is the deposition temperature, and C is the concentration of precursor in the reactor.

Table 1. Constants for the Film Growth Case Study

Constant	Value	Units	Description
κ	0.15	—	Monolayers[23]
ρ	2	—	Number of adatoms per stable cluster
R	8.3145e-3	$\frac{kJ}{mol^*K}$	Gas constant
C_{max}	1.5	$\frac{\mu mol}{L}$	Maximum C
T_{max}	1073	K	Maximum T
u	5.061	$\frac{L}{min}$	Flow of inert gas
A_{sub}	5.067e14	nm^2	Area of substrate
Ω	4.488e22	$\frac{nm^3}{mol}$	Molar volume
h_{ML}	0.2651	$\frac{nm}{ML}$	Thickness of a monolayer[24]
E_a	19.25	$\frac{kJ}{mol}$	Activation energy for flux
E_i	0	eV	Binding energy
E_d	0.8	eV	Diffusion activation energy
η	0.2	—	Capture number

The flux of precursor to the substrate is F, and the true flux of the system is simulated using

$$F = F_0 \exp\left(\frac{-E_i}{RT}\right) \tag{12}$$

where F_0 is the flux as $T \to \infty$, and E_a is the activation energy for the flux. F_0 is calculated as

$$F_0 = \frac{C\Omega u}{A_{sub} h_{ML}} \tag{13}$$

where C is the concentration. The other factors and their values are listed in Table 1. F_0 is a function of C, making F a function of T and C. E_a was calculated assuming a flux of 300 $\frac{ML}{min}$,[23] at the maximum temperature $T_{\max} = 1073\ K$, and maximum concentration $C_{max} = 1.5\ \frac{\mu mol}{L}$. Inserting these quantities into Eqs. 12 and 13, one obtains $E_a = 19.25\ \frac{kJ}{mol}$. The value of the constants in Eqs. 12 and 13 can be found in Table 1. The flux is then used to calculate the growth time for a film. This is calculated using a target film thickness of 150 nm divided by the flux in $\frac{nm}{min}$.

The four models we propose will calculate the flux of atoms to the substrate. Gaussian noise with zero mean and standard deviation in Table 2 is added to the flux simulated by Eqs. 12 and 13. The first model is empirical with one fitted parameter

$$\hat{y}_1(x) = \theta_1 \tag{14}$$

which is the average of the data when doing a least-squares fit. The second model is also empirical, but now with two fitted parameters

$$\hat{y}_2(x) = \theta_1 C + \theta_2 \tag{15}$$

which assumes there is no temperature dependence on the flux going to the substrate. The third model is empirical with three fitted parameters

$$\hat{y}_3(x) = \theta_1 + \theta_2 T + \theta_3 C \tag{16}$$

where the model does assume temperature and concentration dependence for flux. The fourth model is mechanistic and is similar to Eq. 12, but with E_a as the fitted parameter and $F_0 = 10C$, calculated as

$$\hat{y}_4(x) = 10C \exp\left(\frac{-\theta_i}{RT}\right) \tag{17}$$

thus, creating model mismatch from the true system.

The nucleation density is given by Eqs. 18 and 19, and is based on mechanistic principles.[23] The set of differential equations are integrated by the Matlab function `ode23t` to predict the final value of the island density N_{isl}, at κ monolayers (ML) of deposition and depending on the value of the flux F

$$\frac{dN_1}{dt} = F(1-\kappa) - (\rho+1)K_{nuc}(\eta, T, E_i, E_d, N_1) - K_{agg}(\eta, T, E_i, E_d) \tag{18}$$

$$\frac{dN_{isl}}{dt} = K_{nuc}(\eta, T, E_i, N_1) \tag{19}$$

where N_1 is the density of isolated atoms on the surface. The flux F, is the output of the four proposed models. K_{nuc} is the nucleation rate, and K_{agg} is the aggregation rate.[23] The rest of the factors are listed in Table 1. There is no mismatch in this portion of the model.

The objective function for the film growth study is

$$min_x(f_{j^*}(x)) = (t(\hat{y}_{j^*}(x)) - t_{goal})^2 + (10^4(N_{isl}(x, \hat{y}_{j^*}(x)) - N_{goal}))^2 \tag{20}$$

N_{goal} is the nucleation density desired for the film, and t_{goal} is the desired growth time for the film. Values for both are shown in Table 2. This objective is chosen since chemical vapor deposition will usually be one of many steps in creating a product and such specifications would be given at the factory level. Thus, N_{goal} is a target on the quality, while t_{goal} is a target for the process. The confidence interval for the objective function is calculated using the confidence interval on the $t(\hat{y}_{j^*}(x))$ and $N_{isl}(\hat{y}_{j^*}(x))$ predictions based on the confidence interval for F. This is accomplished using absolute errors expanded to second-order terms,[25] due to the

Table 2. Parameters for the Simulated Experimental Data for Nucleation Study in Eqs. 18 and 19

Parameter	Value	Description
v_e	1	Repetitions
γ	5, 10, 20 $\frac{ML}{min}$	Gaussian standard deviation of flux
ε_{CI}	1 $\frac{ML}{min}$	Desired confidence interval for nucleation study
ε_{tol}	0.1	Tolerance for change in MSE
t_{goal}	3 minutes	Growth time objective
N_{goal}	$10^{-3}\frac{islands}{adsorption\ site}$	Nucleation density objective
T	873K	Low T setting
	1073K	High T setting
C	0.3 $\frac{\mu mol}{L}$	Low C setting
	1.5 $\frac{\mu mol}{L}$	High C setting

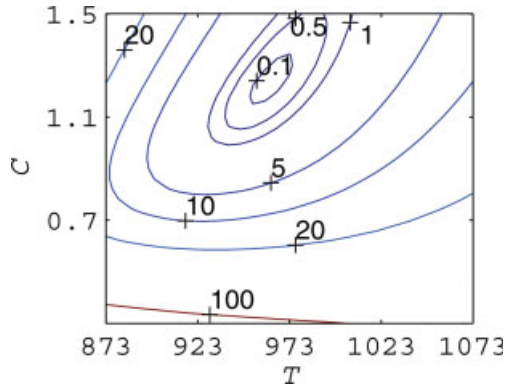

Figure 4. Contour plot of $f(x)$ for the film growth case study with no added noise

[Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

zero slope at the optima. Figure 4 is a contour plot of the objective function calculated using Eq. 20.

The simulated experimental data are generated using Eqs. 12, 13, 18, and 19, using the parameters in Table 2. The study has zero mean Gaussian random noise (γ) added to the data, with the standard deviation given in Table 2. The initial simulated data are generated from a 2^2. factorial experiment using the high-and low-settings for T and C given in Table 2, with one repetition at each design point for a total of four experiments.

Results

All simulations were performed using Matlab Release 2007a, and all optimizations were carried out using Matlab's function fmincon for constrained optimization, which uses an SQP method. The initial guess for finding the predicted optimal point is found by evaluating the model at each of the 256 grid points over the experimental design region. The minimum of these points is chosen as the initial guess for finding x^*.

MHF case study

Typical experimental design surfaces with respect to the parameters x_1 and x_2 are shown in Figure 5a and b. The values on the vertical axis of the D-optimal surface are negative due to a negative sign in front of the optimality calculation,

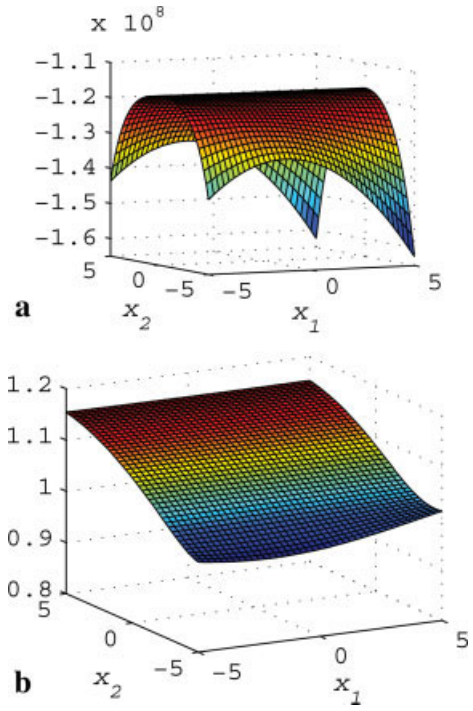

Figure 5. Values of optimality functions in the MHF case study (γ = 3, first iteration) vs. parameters x_1 and x_2. θ = [1.47 19.33 226.18].

(a) D-optimal vs. parameters for the MHF case study after the first iteration, and (b) P-optimal vs. parameters for the MHF case study after the first iteration. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

so that the function can be minimized instead of maximized. The D-optimal surface shows that the corners of the design space are all local minima for this design function. When designing an experiment using D-optimal and without the grid algorithm, experiments are designed only at these corner points, even though these points have already been run, or the corner points are not near the optimal point, or both. The P-optimal surface, on the other hand, has its minima not on the corners of the design space, but along the $x_2 = -5$ edge of the design space. Like D-optimal, P-optimal also tends to prescribe experiments on the boundaries, but the direction is influenced by the value of $\hat{x}$. Recall that in this case study $\bar{x} = [-3.80 - 3.32]$.

Table 3. MHF Case Study with 2^2 + 2 × (0,0) Initial Experiments Using No Grid, ε_{tol} = 0.1

	γ=3			γ=15			γ=30		
	D	P	Rand	D	P	Rand	D	P	Rand
$\hat{y}_{j^*}(\hat{x}_{j^*})$	14.7	14.7	16.4	15.1	15.4	16.5	15.2	15.5	12.0
$\sigma_{j^*}(\hat{x}_{j^*})^2$	1.8	1.2	65.5	35.5	26.6	138	153	94.0	278
$CI(\hat{x}_{j^*})$	2.9	2.4	16.5	12.2	10.7	24.0	25.4	20.0	34.3
ζ (%)	66	66	62	65	64	62	65	64	72
MSE_{j^*}	5.7	5.3	285	170	172	558	692	641	1080
$\hat{\sigma}^2_{j^*}$	8.9	8.3	375	234	244	741	963	894	1480
iter	2.3	2.1	6.1	5.3	4.7	6.1	5.2	4.8	5.3
x_1, % error	0.8	0.8	0.7	0.8	0.7	0.3	0.1	0.6	0.8
x_2, % error	4.1	4.1	4.4	4.1	4.0	4.3	9.6	5.7	17.0

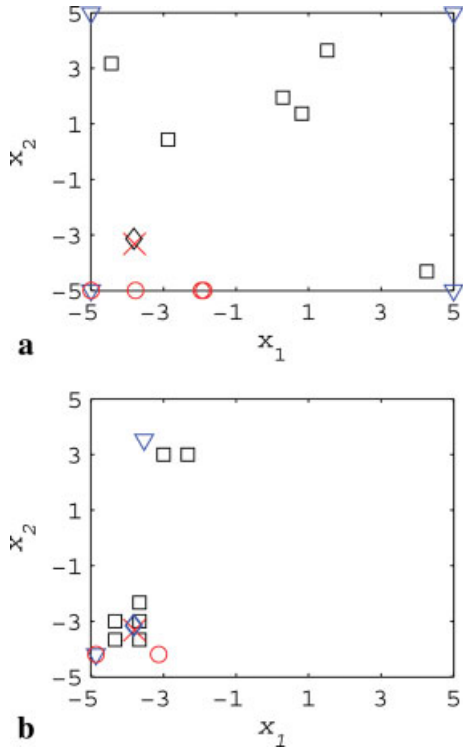

Figure 6. Experiments from the different experimental design methods are marked: D-optimal (▽), P-optimal (o), Random (□), and Greedy (◇).

The greedy experimental design samples at $\hat{x}_{j^*}$ only. × marks the true optimal point. (a) MHF case study experiments without grid algorithm, and (b) MHF case study experiments with grid algorithm. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

The results are summarized in Tables 3–6. These tables are broken up into three different columns for each level of noise. Each of these columns is further split into three columns, one for each type of sequential experimental design function considered. The Rand columns are the results from sequential experiments sampled randomly over the design space. Each table has nine rows which give the final values after the algorithm and experiments have terminated. The first row is $\hat{y}_{j^*}(\hat{x}_{j^*})$, from the most probable model j^* evaluated at $\hat{x}_{j^*}$. The second row is the prediction variance, and the third row is the confidence interval, both at $\hat{x}_{j^*}$. The fourth row is the percentage error between $\hat{y}_{j^*}(\hat{x}_{j^*})$ and $y^{MHF}(\bar{x})$, which is represented by ζ. The fifth row is the mean-squared error as calculated in Eq. 2. The sixth row is the estimated model variance as calculated in Eq. 6, and the seventh row is the average number of iterations. The eighth and ninth row contain the percentage error in $\hat{x}_{j^*}$ compared to $\bar{x}$. For this case study, $\hat{x}$ never matches $\bar{x} = [-3.80 - 3.32]$ exactly, which is the effect of the model mismatch. With some of the terms missing from the model in Eq. 11, compared to Eq. 10, the model's minima may not ever go through the true optimal point, no matter what experiments are performed. Throughout this case study, only designs at low-noise are able to meet the goal of $CI(\hat{x}_{j^*}) < 5$, while most stop because the MSE stopped changing within $\varepsilon_{tol} = 10\%$.

In Table 3, each column is an average of 100 simulations. One can see that without the grid algorithm, none of the experimental design methods leads to good performance at any noise level. The error in $\hat{y}_{j^*}(\hat{x}_{j^*})$ is greater than 60% in all noise levels. Another problem is at low γ, the *CI* for D- and P-optimal are extremely misleading. With such a small *CI*, an experimenter may think his model prediction is very good. This small *CI* can be traced back to the small *MSE*. Figure 6 shows experiments from different experimental design methods after the initial experiments have been run. Comparing Figures 3 and 6a, D- and P-optimal sample along the edges of the design space without the grid algorithm, where $y^{MHF}(x)$ has very different values than $y^{MHF}(\bar{x})$. This leads to a poor fit at $\hat{y}_{j^*}(\hat{x}_{j^*})$, but a small *MSE* based on the sampled points.

Wanting to sample closer to $\bar{x}$, we implement the grid algorithm. In Figure 6b, one can see that the grid algorithm does cause sampling closer to $\bar{x}$. The experiments that are not near $\bar{x}$ are instead close to another local minimum. Strength of the grid algorithm is the ability to search other potential local optima to find the best optimum in the experimental region. If the initial experiments lead to an incorrect $\hat{x}$, the method will not prohibit sampling at other potential optima.

In Table 4, one can see that the grid algorithm does lead to an improvement ($\zeta \approx 40\%$ in the low-noise case instead of ≈60%), although the results still are not very good. Surprisingly, the random experiments with the grid method seem to lead to the best predictions of $\hat{y}_{j^*}(\hat{x}_{j^*})$. In this case, the grid algorithm identifies the potential optima, and the next experiment is chosen at random from these points. D- and P-optimal, on the other hand, use the potential optima from the

Table 4. MHF Case Study with $2^2 + 2 \times (0,0)$ Initial Experiments Using $\delta = 15$, $\varepsilon_{tol} = 0.1$

	$\gamma = 3$			$\gamma = 15$			$\gamma = 30$		
	D	P	Rand	D	P	Rand	D	P	Rand
$\hat{y}_{j^*}(\hat{x}_{j^*})$	25.8	24.9	27.9	24.0	24.4	30.6	19.3	21.6	24.2
$\sigma_{j^*}(\hat{x}_{j^*})^2$	19.9	14.2	50.6	46.3	31.1	83.3	151	112	235
$CI(\hat{x}_{j^*})$	9.2	7.9	14.6	14.0	11.5	18.8	25.4	21.8	31.5
ζ (%)	40	42	36	45	44	29	55	50	44
MSE_{j^*}	113	77.7	250	231	205	376	727	736	955
$\hat{\sigma}^2_{j^*}$	154	113	337	322	287	519	1020	1030	1320
iter	5.3	3.8	5.8	4.9	4.8	5.2	4.7	4.8	5.2
x_1, % error	0.6	0.6	0.6	0.6	0.6	0.6	0.0	0.2	1.1
x_2, % error	4.6	4.8	4.8	4.5	4.8	4.7	9.7	8.1	19.1

Table 5. MHF Case Study with Latin Hypercube Initial Experiments Using $\delta = 15$, $\varepsilon_{tol} = 0.1$

	$\gamma = 3$			$\gamma = 15$			$\gamma = 30$		
	D	P	Rand	D	P	Rand	D	P	Rand
$\hat{y}_{j^*}(\hat{x}_{j^*})$	18.6	20.2	27.0	17.2	19.4	26.1	15.1	19.2	22.8
$\sigma_{j^*}(\hat{x}_{j^*})^2$	106	69.2	219	153	108	382	220	201	488
$CI(\hat{x}_{j^*})$	21.5	17.3	30.6	25.7	21.6	40.5	30.7	29.4	45.7
ζ (%)	57	53	37	60	55	40	65	56	47
MSE_{j^*}	429	382	572	623	577	799	962	1030	1130
$\hat{\sigma}^2_{j^*}$	622	549	820	908	836	1150	1370	1470	1590
iter	4.0	4.0	4.4	3.9	4.0	4.3	4.4	4.3	4.6
x_1, % error	0.3	1.2	2.4	0.1	1.0	4.5	1.5	2.8	7.2
x_2, % error	5.3	16.7	23.1	8.8	14.8	27.1	19.6	29.2	30.6

grid algorithm, and then design the experiment at the point with the best respective value of the design function. Experimental designs like D- and P-optimal tend to maximize the variance of x, which becomes counterproductive when we are dealing with partial models. Unless the experiments are designed to sample at $\hat{x}$, $\hat{x}$ is usually not the point chosen by the experimental design for the next experiment. The random method also samples at a larger variety of experimental points around the optima, whereas the D- and P-optimal methods converge to the same points in subsequent iterations. Note we are not prohibiting any of the designs from repeating previous experiments, and the figures do not explicitly indicate repeated experiments. If one designs the D- and P-optimal algorithms to always pick different points, the predictions from those optimality's might be improved.

Since the random experiments work better with the grid algorithm, and we would like to improve the predictions of $\hat{y}_{j^*}(\hat{x}_{j^*})$ even further, we investigate whether changing our initial experiments to a more random design will improve the results. The initial experiments are run using a Latin Hypercube design as the initial set of six experiments. Comparing Tables 4 and 5 one can see that randomizing the initial experiments does not have the desired effect of improving $\hat{y}_{j^*}(\hat{x}_{j^*})$, although one does obtain results in fewer iterations. Randomizing the initial experiments has the effect of being too space-filling. The *CI* are larger and more accurate, but this has the effect of including too many potential optima in the grid, effectively expanding the sampling area which the grid algorithm is designed to reduce. One also sees increased $\bar{x}$ error, especially in x_2. If one refers to Eq. 11, one sees that x_1 and x_2 are not completely independent, which may be why the error is concentrated on x_2, and not distributed equally between both. Since the LHS method does not provide a significant benefit, we proceed with our initial $2^2 + 2\times(0,0)$ initial design.

One of the parameters of the grid algorithm is the number of points on the grid of the design space, δ, which could be varied to improve $\hat{y}_{j^*}(\hat{x}_{j^*})$. The simulations are run for $\delta = 5$ (coarse grid), and $\delta = 30$ (fine grid). Due to space the detailed results are not included. Overall, we find that a coarse grid had the effect of poorer model predictions than $\delta = 15$. With a coarse grid and high-noise, the constrained search area may also contain many local optima, which hinders the search for the true optimal point. A fine grid, $\delta = 30$, leads to improved predictions, but takes significantly more computational time. For this work, $\delta = 15$ is deemed a good compromise.

A lowered ε_{tol} leads to better predictions from the model and tighter *CI* when comparing to $\varepsilon_{tol} = 0.1$ to $\varepsilon_{tol} = 0.05$, but also more iterations, as one might expect. If the objective of the experiments is to get an accurate model prediction, a smaller tolerance (i.e., willingness to run more experiments) will cause a better fit. However, if the emphasis is on quickly finding the optimal point, then a larger ε_{tol} may produce adequate results. At any point during the design phase, one may choose to terminate the experiments, before reaching the prespecified tolerance.

Since sampling near the optimum point seems to improve $\hat{y}_{j^*}(\hat{x}_{j^*})$, why not sample at the predicted optimum point? We implemented this using a "greedy" method, and the results are shown in Table 6, and marked by a $\diamond$ in Figure 6b. Rather than minimize a separate experimental design function, this method chooses the estimated optimal point of the model j^* as the next experimental point to be sampled. This leads to similar predictions and number of iterations to the random plus grid algorithm for low-noise ($\delta = 30$, $\varepsilon_{tol} = 0.1$), while for high-noise the performance and iterations are similar to random plus grid algorithm with $\delta = 15$ and $\varepsilon_{tol} = 0.05$. These similarities point to the importance of the stopping criterion for the experimental design. The grid algorithm can obtain similar results to the greedy method given the right stopping criterion, but has the advantage of exploring more of the experimental region than if one were to use the greedy method. The better model prediction of greedy makes sense as more experiments are run and experiments are located at the optimal point. The least-squares method will then focus on minimizing the error from these points, making the fit to $y(\bar{x})$ better. This "greedy" case can be viewed as the limit of the grid algorithm with $\delta\to\infty$ and $\alpha\to100\%$. Ultimately, the performance in this case study is limited by the model mismatch between Eqs. 10 and 11. One could

Table 6. MHF Case Study With with $2^2 + 2 \times (0,0)$ Initial Experiments and "Greedy" Experimental design, $\varepsilon_{tol} = 0.1$

	$\gamma = 3$	$\gamma = 15$	$\gamma = 30$
$\hat{y}_{j^*}(\hat{x}_{j^*})$	33.6	37.0	35.7
$\sigma_{j^*}(\hat{x}_{j^*})^2$	24.0	37.3	111
$CI(\hat{x}_{j^*})$	10.0	12.3	21.2
ζ (%)	22.5	14.7	17.7
MSE_{j^*}	168	342	940
$\hat{\sigma}^2_{j^*}$	225	427	1180
iter	6.1	10.3	10.0
x_1, % error	0.48	0.01	1.48
x_2, % error	5.25	9.06	20.3

Table 7. Film Growth Case Study with 2^2 Initial Experiments Using No Grid, $\varepsilon_{tol} = 0.1$

	$\gamma = 5$			$\gamma = 10$			$\gamma = 20$		
	D	P	Rand	D	P	Rand	D	P	Rand
$f_{j^*}(\hat{x})$	0.05	9×10^{-9}	0.02	0.07	5×10^{-8}	0.03	0.02	0.01	0.04
$\sigma_{j^*}(\hat{x}_{j^*})^2$	221	10.8	26.8	304	141	108	754	637	515
$CI(\hat{x}_{j^*})$	7.40	1.95	2.80	9.03	5.24	5.46	16.1	13.6	9.93
MSE_{j^*}	169	7.1	8.48	255	125	33.6	676	579	138
$\hat{\sigma}^2_{j^*}$	675	28.4	30.6	907	401	115	2190	1800	434
iter	8.1	6.7	8.8	8.4	7.2	9.0	7.7	7.4	9.4
$\hat{T}$, % error	1.54	0.47	0.02	2.91	0.46	0.2	1.43	0.05	0.03
$\hat{C}$, % error	9.19	2.85	0.19	17.25	2.82	1.49	8.70	0.35	0.03

probably increase the prediction accuracy more by leaving out the initial experiments during the parameter fitting, which would cause the fit to be better around the more recently sampled points.

Film growth case study

The film growth case study is different from the first case study because four candidate models are now used. The first case study always underpredicted the optimum, but in this case study the models are able to predict the desired output of the process ($t = 3$ min and $N_{isl} = 10^{-3}$); however $\hat{x}_{j^*}$ is not necessarily $\bar{x}$. For this reason we restricted the models to only fit data that falls below $H(\hat{x})$, or at least enough data to make A be full rank. This will increase the confidence interval on the prediction due to fitting to fewer points, but ultimately should enable more accurate fits at the optimal point. The true optimum in this case is at $\bar{T} = 962\,K$ and $\bar{C} = 1.24\frac{\mu mol}{\min}$. The tables in this section show the objective function for the most probable model, instead of the output, $y(x)$, since $f(x) \neq y(x)$. The $CI(\hat{x}_{j^*})$ shown is also for $f(x)$ and not on $y(x)$ for the same reason.

In Table 7, the third empirical model (Eq. 16) is the most probable in two thirds of the simulations, the second empirical model (Eq. 15) or the mechanistic model (Eq. 17) were the next most probable in equal proportion, and the first empirical model is picked as most probable the least often of the four. Here, $f_{j^*}(\hat{x})$ are quite small ($f_{j^*}(\hat{x}) < 0.1$), and the $CI(\hat{x}_{j^*})$ are quite large compared to the size of the objective function at the optimal point. In the no grid case, all three experimental designs predict that $f_{j^*}(\hat{x}) \approx 0$ at all noise levels. The random experimental design has the smallest error for predicting $\hat{T}$ and $\hat{C}$. In Figure 7a, one can see the experiments for each experimental design. The experiments for D- and P-optimal are on the corners of the design region, far away from the optimal point. The random design samples all over the design region leading to the better $\hat{x}$ prediction.

In Table 8 the results of the film growth experiments using the grid algorithm are shown. At first glance, one would say that the grid did not bring any improvement to the experiments at all. In fact, the $CI(\hat{x}_{j^*})$ are not significantly different than without the grid. However, the % error in $\hat{T}$ and $\hat{C}$ is improved with the grid for D- and P-optimal. One can see the effect of the grid on the experiments in Figure 7b, where there are more experiments around the optimal point than without the grid. In Figure 8 the possible optimal grid points are plotted for sequential experiments 1, 4, and 7 for one simulation of the random experimental design. At each successive experiment, the range of possible optimal points shrinks. It is also worth noting that in iterations 1 and 4, the grid retains the lower left corner of the experimental area. Here, the confidence interval for these points is so large that they are included as possible optima.

One interesting effect of the grid algorithm is in the resulting most probable model for each experimental design. Using the grid algorithm, the most probable model is still the third empirical model, but the mechanistic model is the second most likely, and the first empirical model is never considered

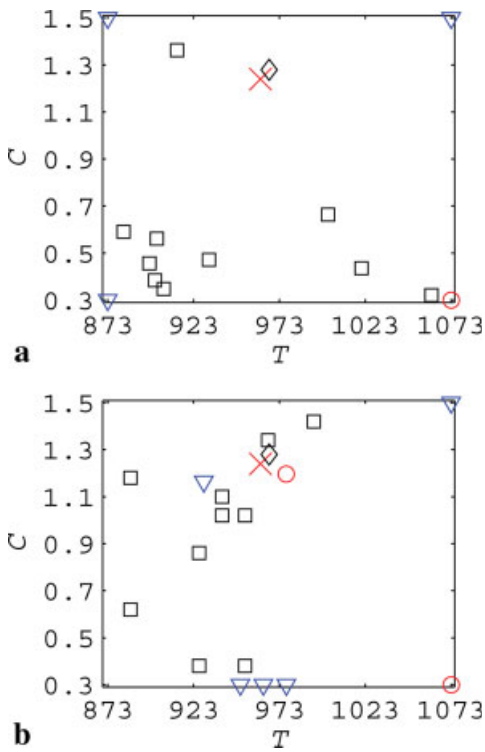

Figure 7. Experiments from the different experimental design methods are marked: D-optimal (▽), P-optimal (○), Random (□), and greedy (◇).

The greedy experimental design samples $\hat{x}_{j^*}$ at only. × marks the true optimal point, with $\gamma = 10$ s, and $\delta = 15$. (a) Film growth case study experiments without grid algorithm, and (b) film growth case study experiments with grid algorithm. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com

Table 8. Film Growth Case Study with 2^2 Initial Experiments Using $\delta = 15$, $\varepsilon_{tol} = 0.1$

	$\gamma = 5$			$\gamma = 10$			$\gamma = 20$		
	D	P	Rand	D	P	Rand	D	P	Rand
$f_{j^*}(\hat{x})$	0.12	2×10^{-5}	2×10^{-8}	0.03	2×10^{-3}	2×10^{-8}	0.03	6×10^{-4}	2×10^{-3}
$\hat{f}_{j^*}(\hat{x})$	392	106	21.3	331	150	67.6	885	738	845
$\sigma_{j^*}(\hat{x}_{j^*})^2$	16.1	3.98	2.64	11.9	6.11	4.45	18.0	15.4	10.4
MSE_{j^*}	227	65.8	12.5	248	130	37.3	774	677	133
$\hat{\sigma}^2_{j^*}$	933	223	43.1	894	436	120	2440	2110	396
iter	7.4	7.4	8.4	7.3	7.5	9.0	7.5	7.5	9.2
x_1, % error	0.03	0.15	0.14	0.31	0.30	0.05	1.19	0.07	0.03
$\hat{C}$, % error	0.36	0.88	1.53	2.13	1.81	0.77	7.47	0.70	0.41

the most probable model as it sometimes was without the grid. The grid had the effect of eliminating one of the possible models, and clearly identified a temperature dependence in the data (most probable models both included T) that was not apparent without the grid. The mechanistic model does well with better sampling of the experimental area around the optimal point, but still not better than Eq. 16. This indicates that the mechanistic model may need modification for improved prediction, but not necessarily that a whole new model is needed to explain the data.

The film growth experiments were performed using the greedy experimental design where only the optimal point is used as the next experiment, but the results are not shown here due to space. In this case study, the greedy experimental design does not yield significant improvements to finding the objective function C1 or the optimal point. However, the greedy experimental design picks the mechanistic model as the most probable model rather than any of the empirical models. Again, the greedy method does do quite well in most simulations, but in some simulations $\hat{x}$ is incorrect, and sampling at the same incorrect point repeatedly does not improve $\hat{x}$. Good performance is achieved using either the grid algorithm or the random sampling in this case study. Based on both case studies, combining the two together will provide good estimates of the optimal point, while still allowing for some exploration and model building.

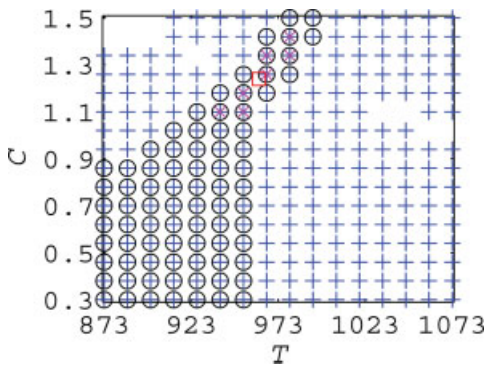

Figure 8. Possible optimal points after grid algorithm using the random experimental design.

First sequential experiment marked by (+), fourth sequential experiment marked by (○), and seventh seqential experiment marked by ×. The true optimal point is [962 1.24], and is marked by (□). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Conclusions

The grid method in the MHF case study used in conjunction with other experimental designs results in enhanced performance of these designs. Improvements such as better model predictions (15–20% better than without the grid algorithm), and more accurate confidence intervals which contain the true optimum of the MHF model (43.3) are observed here. The film growth case study achieved good results using random experimental design; however, the grid algorithm did lead to better sampling of the experimental region, better $\hat{x}$ prediction, and led the experimenter to a dependence on T that would have been missed otherwise. The P-optimal design was introduced, but due to model mismatch did not produce the desired effect of improving the prediction at the optimal point. Random experimental design used in conjunction with the grid algorithm worked best.

In this work, experimental design and its effect on model building was discussed. Parameter estimation also plays a pivotal role in model building which is an opportunity for future study. As with any experimental design method, it is up to the experimenter to define the objective for the experiment and to interpret the results. In this work, the main objective is to obtain good model predictions of $\bar{x}$ of a process for use in process design. The grid algorithm is a useful tool because it balances the trade-off between an optimal design such as D-optimal and a "greedy" experimental design. It also allows for exploring other regions with potential optima rather than focusing on one optima which may be erroneous. The methodology presented here combines existing experimental design approaches for empirical models with those for mechanistic models. Thus, the experimenter does not have to decide ahead of time whether to use empirical only, or mechanistic only, since both may be useful. As with many experimental design techniques, there is not one setting for the grid algorithm or stopping criteria that works for all experiments. The experimenter needs to combine his knowledge of the system being studied and interpretation of the available data to come up with a viable experimental design.

Notation

C = concentration of precursor in gas
$CI_j(x)$ = confidence interval on j^{th} model
E_d = diffusion activation energy
E_i = binding energy

$f(x)$ = objective function
F = flux
$H(x)$ = threshold value for grid algorithm
j^* = most probable model among a set of j possible models
K_{agg} = aggregation rate
K_{nuc} = nucleation rate
MSE_j = mean-square error of model j
n = number of total experiments run
N_1 = isolated atom ("adatom")
N_{isl} = nucleation density on surface of substrate
N_{goal} = desired nucleation density
p_j = number of parameters in model j
$P(Mj)$ = probability of model j being the correct model
$t(\hat{y}_{j^*}(x))$ = time needed to grow a film of a certain thickness from model j^*
T = temperature setting for film growth case study
$\bar{x}$ = true optimal point of process
$\hat{x}$ = estimated optimal point
$y(x_i)$ = the i^{th} experimental data
$\hat{y}_j(x)$ = prediction from the j^{th} model
Y = experimental data
δ = number of points used in grid algorithm
ε_{CI} = desired confidence interval on best model
ε_{tol} = maximum allowed change in MSE for stopping criterion
ς = % error in $y_{j^*}(\hat{x}_{j^*})$
η = capture number
θ_j = parameter set for j^{th} model
κ = number of monolayers of deposition
v_e = number of repetitions of experiments
ρ = number of adatoms needed for a stable cluster
$\hat{\sigma}_j^2$ = model variance
$\sigma_j^2(x)$ = prediction variance at point x

Literature Cited

1. Topol AW, Dunn KA, Barth KW, Nuesca GM, Taylor BK, Dovidenko K, Kaloyeros AE, Tuenge RT, King CN. Chemical vapor deposition of ZnS : Mn for thin-film electroluminescent display applications. *J Mater Res*. 2004;19(3):697–706.
2. Rawlings JB, Miller SM, Witkowski WR. Model identification and control of solution crystallization processes: a review. *Ind Eng Chem Res*. 1993;32(7):1275–1296.
3. Braake HAB, van Can HJL, Verbruggen HB. Semi-mechanistic modeling of chemical processes with neural networks. *Eng Appl Artif Intel*. 1998;11(4):507–515.
4. Tulleken HJAF. Grey-box modelling and identification using physical knowledge and Bayesian techniques. *Automatica*. 1993;29(2): 285–308.
5. Stewart WE, Shon Y, Box GEP. Discrimination and goodness of fit of multiresponse mechanistic models. *AIChE J*. 1998;44(6):1404–1412.
6. Froment GF. Model discrimination and parameter estimation in heterogeneous catalysis. *AIChE J*. 1975;21(6):1041–1057.
7. Qian ZG, Seepersad CC, Joseph VR, Allen JK, Wu CFJ. Building surrogate models based on detailed and approximate simulations. *J Mech Des*. 2006;128(4):668–677.
8. Weiss L, Amon C, Finger S, Miller E, Romero D, Verdinelli I, Walker L, Campbell P. Bayesian computer-aided experimental design of heterogeneous scaffolds for tissue engineering. *Comp-Aided Des*. 2005;37(11):1127–1139.
9. Fujiwara M, Nagy ZK, Chew JW, Braatz RD. First-principles and direct design approaches for the control of pharmaceutical crystallization. *J Process Contr*. 2005;15(5):493–504.
10. Studden WJ. Ds-optimal designs for polynomial regression using continued fractions. *Annals Stats*. 1980;8(5):1132–1141.
11. Franceschini G, Macchietto S. Validation of a model for biodiesel production through model-based experiment design. *Ind Eng Chem Res*. 2007;46(1):220–232.
12. Montgomery DC. *Design and Analysis of Experiments. vol. 6*. John Wiley & Sons, Inc; 2005.
13. Myers RH, Montgomery DC. *Response Surface Methodology: Process and Product Optimization using Designed Experiments*. New York: Wiley; 1995.
14. Phadke M. *Quality Engineering Using Robust Design*. P T R Prentice-Hall, Inc; 1989.
15. Diwekar UM, Rubin ES. Parameter design methodology for chemical processes using a simulator. *Ind Eng Chem Res*. 1994;33(2):292–298.
16. Stein A, Ettema C. An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons. *Agric Eco Environ*. 2003;94(1):31–47.
17. Box GEP, Hill WJ. Discrimination among mechanistic models. *Technometrics*. 1967;9(1):57.
18. Kirkpatrick S, Gelatt C, Vecchi M. Optimization by simulated annealing. *Science*. 1983;220:671–680.
19. Knowles J, Corne D. A new evolutionary approach to the degree-constrained minimum spanning tree problem. *IEEE Trans Evol Comput*. 2000;4(2):125–134.
20. Xiong RT, Wissmann PJ, Gallivan MA. An extended Kalman filter for in situ sensing of yttria-stabilized zirconia in chemical vapor deposition. *Comp Chem Eng*. 2006;30(10–12):1657–1669.
21. Chen JH, Wong DSH, Jang SS, Yang SL. Product and process development using artificial neural-network model and information analysis. *AIChE J*. 1998;44(4):876–887.
22. Choy KL. Chemical vapour deposition of coatings. *Prog Mater Sci*. 2003;48(2):57–170.
23. JW Evans PT, Bartelt M. Morphological evolution during epitaxial thin film growth: Formation of 2D islands and 3D mounds. *Surf Sci Rep*. 2006;61:1–128.
24. Xu YN, Gu ZQ, Ching WY. Electronic, structural, and optical properties of crystalline yttria. *Phys Rev B*. 1997;56(23):14993–15000.
25. Chapra SC, Canale RP. *Numerical Methods for Engineers*. 4th ed. McGraw-Hill; 2002.

Manuscript received May 23, 2008, and revision received Sept. 8, 2008.