Proceedings of the 9th International Symposium on
Dynamics and Control of Process Systems (DYCOPS 2010),
Leuven, Belgium, July 5-7, 2010
Mayuresh Kothare, Moses Tade, Alain Vande Wouwer, Ilse Smets (Eds.)

TuMT3.5

Iterative Data-based Modelling and Optimization for Rapid Design of Dynamic Processes

Guoyi Chi, Wenjin Yan, Tao Chen

School of Chemical and Biomedical Engineering, Nanyang Technological University, 62 Nanyang Drive, Singapore 637459, Singapore (e-mail: {chig0002, S080062, chentao}@ntu.edu.sg).

Abstract: We consider an off-line process design problem where the response variable is affected by several factors. We present a data-based modelling approach that iteratively allocates new experimental points, update the model, and search for the optimal process factors. A flexible non-linear modelling technique, the kriging (also known as Gaussian processes), forms the cornerstone of this approach. Kriging model is capable of providing accurate predictive mean and variance, the latter being a quantification of its prediction uncertainty. Therefore, the iterative algorithm is devised by jointly considering two objectives: (i) to search for the best *predicted* response, and (ii) to adequately explore the factor's space so that the predictive uncertainty is small. This method is further extended to consider dynamic processes, i.e. the process factors are time-varying and thus the problem becomes to design a time-dependent trajectory of these factors. The proposed approach has been demonstrated by its application to a simulated chemical process with promising results being achieved.

Keywords: Batch Processes, design of experiments, Gaussian process model, kriging model, process optimization, response surface methodology.

1. INTRODUCTION

Mathematical models are the foundation of the systems approach to the design of chemical and other processes (Klatt and Marquardt, 2009). Models can be developed through the representation of fundamental principles that govern the process, and thus they are termed *first-principles* or *mechanistic* models. Alternatively, models may be purely based on experimental data and are called *data-based* or *empirical*. Although data-based models are typically reliable only within the operating region where the data are collected, they have seen wide applications due to the simplicity of model development and implementation. This is especially true if the process is still in its early design stage, whereby the time and resources needed for mechanistic modelling are hardly justifiable.

This study is further restricted to batch-wise (as opposed to time-dependent) modelling that relates the process response (y, e.g. product yield) to the operating factors ($\mathbf{x} = [x_1, \ldots, x_d]^{\mathrm{T}}$, e.g. reaction temperature and pressure), where d is the number of factors. These models are typically used in off-line design stage to facilitate the understanding and optimization of the process. This is in contrast to dynamic process models which are mainly used for on-line control and optimization purposes. Such a data-based model is the central component of the so-called response surface methodology (RSM) for rational process design (Myers and Montgomery, 1995).

The traditional method in RSM is to fit a polynomial function (typically linear, quadratic or cubic polynomial) to the experimental data, followed by identifying the process factors that optimize the objective function. However, the prediction accuracy of the polynomial regression is usually unsatisfactory due to the restrictive functional form, and consequently the model-based process understanding and optimization may be unreliable. To address this issue, flexible non-linear models have been applied to provide a more accurate approximation of the process behaviour, such as artificial neural network (ANN) (Dutta et al., 2004; Shao et al., 2007), support vector machine (SVM) (Hadjmohammadi and Kamel, 2008), and kriging models (also known as Gaussian process regression) (Yuan et al., 2008; Tang et al., 2010). Kriging is particularly attractive since it not only attains high prediction accuracy, but also quantifies the uncertainty of the prediction. Proper handling of the prediction uncertainty is necessary to ensure reliable optimization results.

This paper extends the previous kriging-based process optimization (Yuan et al., 2008; Tang et al., 2010) to a fully iterative approach. Intuitively, two objectives should be considered when allocating new experimental points based on the current model: (i) to search for the best *predicted* response, and (ii) to adequately explore the factor's space. Usually, a large predictive uncertainty within certain region of the factor's space is an indication that this region is not well explored. We adopt a formal statistical framework due to Jones (2001) to jointly account for these two goals. We further extend this approach to the design of time-varying process factors, such as temperature profile. The proposed method is successfully applied to a simulated batch chemical process.
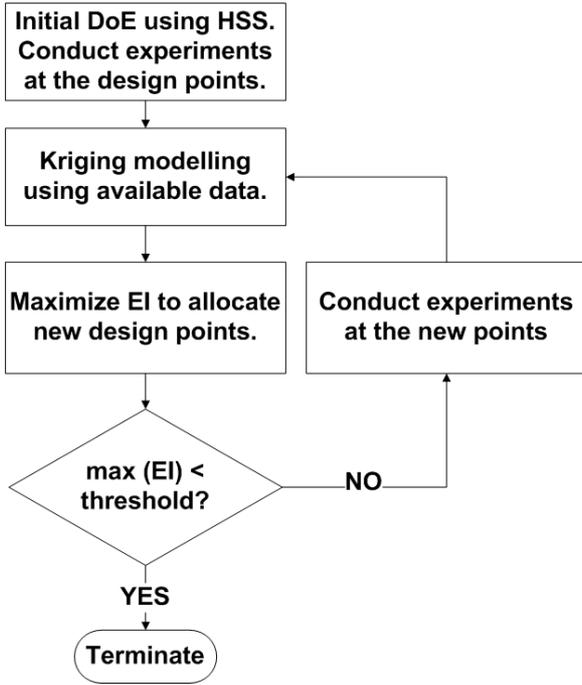
Fig. 1. The flowchart of the proposed algorithm.

2. ITERATIVE MODELLING AND OPTIMIZATION

The overall approach is illustrated in Fig. 1. Initially, statistical design of experiments (DoE) is applied to give the initial design points for conducting experiments. The experimental data are used to develop a kriging model to relate the process response to the factors. Based on the model, we allocate new design points by maximizing the *expected improvement* (to be discussed subsequently), and new experiments will be conducted to update the model. We discuss the various components of this algorithm in this section.

2.1 Initial experimental design

Initial design of experiments (DoE) is required to obtain the data for empirical modelling. The classical fractional factorial and central composite designs were proposed to investigate the interactions of process factors based on polynomial models (Myers and Montgomery, 1995). These classical designs typically assign two or three predetermined levels for each process factor, and experiments are conducted at the combination of the levels of different factors. Using a small number of levels is especially appealing if the factors' values are difficult to change in practice. However, this strategy may not have an optimal coverage of the design space due to limited levels of the factors being studied, and thus it may result in a less reliable empirical model (Fang et al., 2000). The recognition of this disadvantage of classical DoE methods has motivated the concept of "space-filling" designs that allocate design points to be uniformly distributed within the range of each factor (Fang et al., 2000; McKay et al., 1979; Kalagnanam and Diwekar, 1997). Among this class of designs, the Hammersley sequence sampling (HSS) (Kalagnanam and Diwekar, 1997) has received wide application as a result of its simple implementation and good performance. HSS

is adopted in this study for initial DoE, and it is briefly introduced in this subsection.

The HSS design is based on the fact that any integer n can be written in a radix notation of another integer R as follows (Kalagnanam and Diwekar, 1997):

$$
\begin{aligned}
n &\equiv n_0 n_1 n_2 \cdots n_{m-1} n_m \\
&= n_m + n_{m-1}R + n_{m-2}R^2 + \cdots + n_1 R^{m-1} + n_0 R^m
\end{aligned}
\tag{1}
$$

where m is the integral part of $\log_R n$. A function of n, called inverse radix number, can be constructed by reversing the order of the digits of n and concatenating them behind a decimal point:

$$
\begin{aligned}
\psi_R(n) &= 0.n_m n_{m-1} \cdots n_2 n_1 n_0 \\
&= n_m R^{-1} + n_{m-1}R^{-2} + \cdots + n_1 R^{-m} + n_0 R^{-m-1}
\end{aligned}
\tag{2}
$$

Now we select the first $d-1$ prime numbers as the integer R in eq. (1): $R_1, R_2, \cdots, R_{d-1}$. According to HSS, the n design points, each being a vector of order d, are given by

$$
\mathbf{x}_i = \mathbf{1} - \left[\frac{i}{n}, \psi_{R_1}(n), \psi_{R_2}(n) \cdots, \psi_{R_{d-1}}(n) \right]^T
\tag{3}
$$

where $i = 1, 2, \cdots, n$ and $\mathbf{1}$ is a unity vector.

2.2 Overview of kriging model

Suppose that n experimental runs have been conducted and the data are $\{\mathbf{x}_i, y_i, i = 1, \ldots, n\}$. Kriging is based on a stochastic process model to approximate the response-factor relationship:

$$
y(\mathbf{x}) = \mu + \epsilon(\mathbf{x})
\tag{4}
$$

where μ is an unknown constant. $\epsilon(\mathbf{x})$ is a realization of a Gaussian process having zero mean and covariance between the random variables $\epsilon(\mathbf{x}_i)$ and $\epsilon(\mathbf{x}_j)$ given by

$$
\begin{aligned}
\text{cov}\left[\epsilon(x_i), \epsilon(x_j)\right] &= \sigma^2 R(\mathbf{x}_i, \mathbf{x}_j) \\
&= \sigma^2 \prod_{k=1}^{d} \exp\left(-\theta_k |x_{ik} - x_{jk}|^2\right)
\end{aligned}
\tag{5}
$$

where $\mathbf{x}_i = [x_{i1}, \ldots, x_{id}]'$. The parameters of this model $(\mu, \sigma^2, \boldsymbol{\theta} = \{\theta_1, \ldots, \theta_d\})$ can be estimated by the maximizing the following likelihood function

$$
\begin{aligned}
L(\mu, \sigma^2, \boldsymbol{\theta}) =& (2\pi\sigma^2)^{n/2} |\mathbf{R}|^{1/2} \\
& \cdot \exp\left[-\frac{(\mathbf{y} - \mathbf{1} \cdot \mu)' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1} \cdot \mu)}{2\sigma^2} \right]
\end{aligned}
\tag{6}
$$

where $\mathbf{R}$ is an $n \times n$ matrix of correlation function $R(\cdot, \cdot)$ corresponding to pairs of data points, i.e. $R_{ij} = R(\mathbf{x}_i, \mathbf{x}_j)$ as given in eq. (5). The column vector $\mathbf{y}$ consists of the known output values: $\mathbf{y} = [y_1, \ldots, y_n]'$.

Given estimated parameters $(\hat{\mu}, \hat{\sigma}^2, \hat{\boldsymbol{\theta}})$, the prediction at a new data point $\mathbf{x}_*$ is also normally distributed (Sacks et al., 1989) with mean and variance given by

$$\hat{y}_* = \hat{\mu} + \mathbf{r}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1} \cdot \mu) \tag{7}$$

$$s_*^2 = \hat{\sigma}^2 \left[1 - \mathbf{r}'\mathbf{R}^{-1}\mathbf{r} + \frac{(1 - \mathbf{r}'\mathbf{R}^{-1}\mathbf{r})^2}{\mathbf{1}'\mathbf{R}^{-1}\mathbf{1}} \right] \tag{8}$$

where

$$\mathbf{r} = [R(\mathbf{x}_*, \mathbf{x}_1), \dots, R(\mathbf{x}_*, \mathbf{x}_n)]' \tag{9}$$

2.3 Iterative modelling and optimization

A kriging model forms the basis of model-based optimization. A simple strategy would be to collect a certain number of experimental data, develop a kriging model, find the optimal process factors, and finally conduct verification experiment(s) at these factors (Yuan et al., 2008). However, it is difficult to determine *a priori* the amount of data that are sufficient to build a reliable model. Hence, an iterative approach, as given in Fig. 1, is more desirable.

A straightforward method is to find an optimum $\mathbf{x}_{\text{new}}$ based on model predictions, and then conduct new experiment at $\mathbf{x}_{\text{new}}$. However, this is not the best approach, since it ignored the predictive uncertainty that is available in kriging model to quantify the mismatch between the model and real process. For example, a data point that is predicted to give inferior response with very high variance may actually result in optimal process. A large predictive variance typically suggests that the experimental data around this point are not sufficient to give a reliable prediction. Therefore, both predictive mean and variance must be jointly considered in the optimization algorithm.

Previously, it was suggested (Apley et al., 2006) to maximize the lower-bound of the response predicted by the model (suppose the objective is maximization):

$$\mathbf{x}_{\text{new}} = \arg\max_{\mathbf{x}_*} (\hat{y}_* - \alpha s_*) \tag{10}$$

where α is a user-set term. For example, setting $\alpha = 1.645$ corresponds to a 95% lower bound, since the prediction is Gaussian distributed. However, this approach does not properly utilize the predictive uncertainty. For example, if s_* is very large due to the lack of experimental data around its neighbourhood, then this point $\mathbf{x}_*$ will not be selected for experimentation. Should it be selected and the uncertainty be reduced, this point could turn out to be highly desired. We later modified the criterion in eq. (10) to partially address this issue using an *ad hoc* solution (Tang et al., 2010).

In the present study, we adopt the concept of "expected improvement" (EI) (Jones, 2001) that has been introduced to quantify the improvement that is expected to obtain from conducting addition experiments at any design points. Formally for maximization problem, the predicted improvement at $\mathbf{x}_*$ is $I(\mathbf{x}_*) = y(\mathbf{x}_*) - f_{\text{best}}$, where f_{best} is the largest response value obtained through experiments so far, and $y(\mathbf{x}_*)$ is the prediction from the kriging model.

Since $y(\mathbf{x}_*)$ is Gaussian distributed with mean $\hat{y}^*$ and variance s_*^2 (eqs. (7)(8)), the improvement is also Gaussian distributed with mean $\hat{y}^* - f_{\text{best}}$ and the same variance. Therefore, the expected improvement at $\mathbf{x}_*$ is given by

$$\text{EI}(\mathbf{x}_*) = E\left[\max\{0, I(\mathbf{x}_*)\}\right] = \int_0^\infty I p(I) dI$$
$$= s_* \left[u\Phi(u) + \phi(u)\right] \tag{11}$$

where $u = (\hat{y}^* - f_{\text{best}})/s_*$. $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution function and density function of standard normal distribution, respectively. EI will increase if the predicted response is greater than f_{best} and/or the predictive variance is large, and thus further experiments should be conducted at this region.

Therefore, instead of optimizing the mean prediction of the response, we search for process factors $\mathbf{x}$ that maximize EI using, for example, sequential quadratic programming or genetic algorithms. If maximal EI becomes close to zero, then the entire optimization procedure can be terminated since we expect no further improvement by conducting additional experiments.

2.4 Design of dynamic processes

Operating a batch process by following a time-varying profile of, e.g. temperature, is not uncommon. For many processes, the non-isothermal operation can achieve better performance in terms of higher yield or shorter batch duration. Another example of time-varying profile is from fed-batch fermentation processes, whereby the changing flow rate of substrate is desired (Georgakis, 2009).

In the literature, two approaches have been suggested for optimizing time-varying process factors. Georgakis (2009) recommended to use basis functions to represent a dynamic profile $u(t)$:

$$u(t) = \sum_{j=1}^m \beta_j \psi_j(t) \tag{12}$$

where t denotes time, $\{\psi_j(t), j = 1, \cdots, m\}$ is a set of pre-selected basis functions (e.g. orthogonal polynomials or wavelets), and β_j's are the coefficients. Hence the design of an optimal $u(t)$ is equivalent to determining the optimal coefficients β_j's. It was demonstrated in (Georgakis, 2009) that for practical situations, orthogonal polynomials up to the second order would have sufficient freedom for design purpose. However, the major difficulty with this approach is the handling of process constraints. Even for simple interval constraints (e.g. temperature profile is between 20 and 50 °C), some *ad hoc* adjustment has to be made to rule out infeasible solutions.

In this work, we adopt the method of orthogonal collocation to handle time-varying factors (Villadsen and Michelsen, 1978). Specifically, we assign collocation points at several pre-selected time instances throughout the batch duration, and then a Lagrangian interpolation polynomial will pass through the desired values of these collocation points. Effectively, the design problem becomes to find the optimal values of these collocation points, and this

Table 1. Parameters for the batch reversible reaction.

Parameter	Value	Parameter	Value Page
k_{10} (1/h)	1.32×10^7	k_{20} (1/h)	5.24×10^{13}
E_1 (cal/mol)	1×10^4	E_2 (cal/mol)	2×10^4
C_{A0} (mol/l)	1	C_{B0} (mol/l)	0
T (°C)	$20 - 50$		

problem can be solved by using the proposed kriging-based framework.

3. CASE STUDY

We consider the optimization of the operation of a batch reactor in which a reversible reaction between A and B takes place:

$$A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} B \qquad (13)$$

It is assumed that both reactions are first-order, The component mass balances are

$$\frac{dC_A}{dt} = -k_1 C_A + k_2 C_B, \quad C_A(0) = C_{A0} \qquad (14)$$

$$\frac{dC_B}{dt} = k_1 C_A - k_2 C_B, \quad C_B(0) = C_{B0} \qquad (15)$$

where the reaction kinetics are given by

$$k_1 = k_{10} \exp\left(\frac{-E_1}{RT}\right), \quad k_2 = k_{20} \exp\left(\frac{-E_2}{RT}\right) \qquad (16)$$

The process is simulated by solving above differential equations with the parameters given in Table 1.

3.1 Maximize the process conversion at a given time

We consider the first scenario to maximize the conversion rate of reactant A at the end of the reaction. For illustration purpose, we set the nominal batch duration to be 2.5 h.

We first consider a isothermal operation to search for the fixed temperature that gives the highest conversion. To initialize, the HSS was used to give three design points, and the process model was solved to obtain the corresponding conversion rates. (In practice, real experiments would be conducted at the designed temperature to get this information.) Then, the proposed algorithm is followed to optimize the conversion iteratively.

Fig. 2 illustrates the kriging model prediction and the corresponding EI at the first three iterations. At the first iteration (Fig. 2(a)), the prediction uncertainty is clearly observed (the shaded area). If we only wish to reduce the uncertainty, we would have allocated the new experimental point around 50 °C. However, by considering the joint effect of uncertainty reduction and process optimization, the criterion of EI allocates the new experimental point at 32.9 °C. By conducting this new experiment, the prediction uncertainty between 20 and 40 °C has been dramatically reduced (Fig. 2(b)), and the new experiment

(a)

(b)

(c)
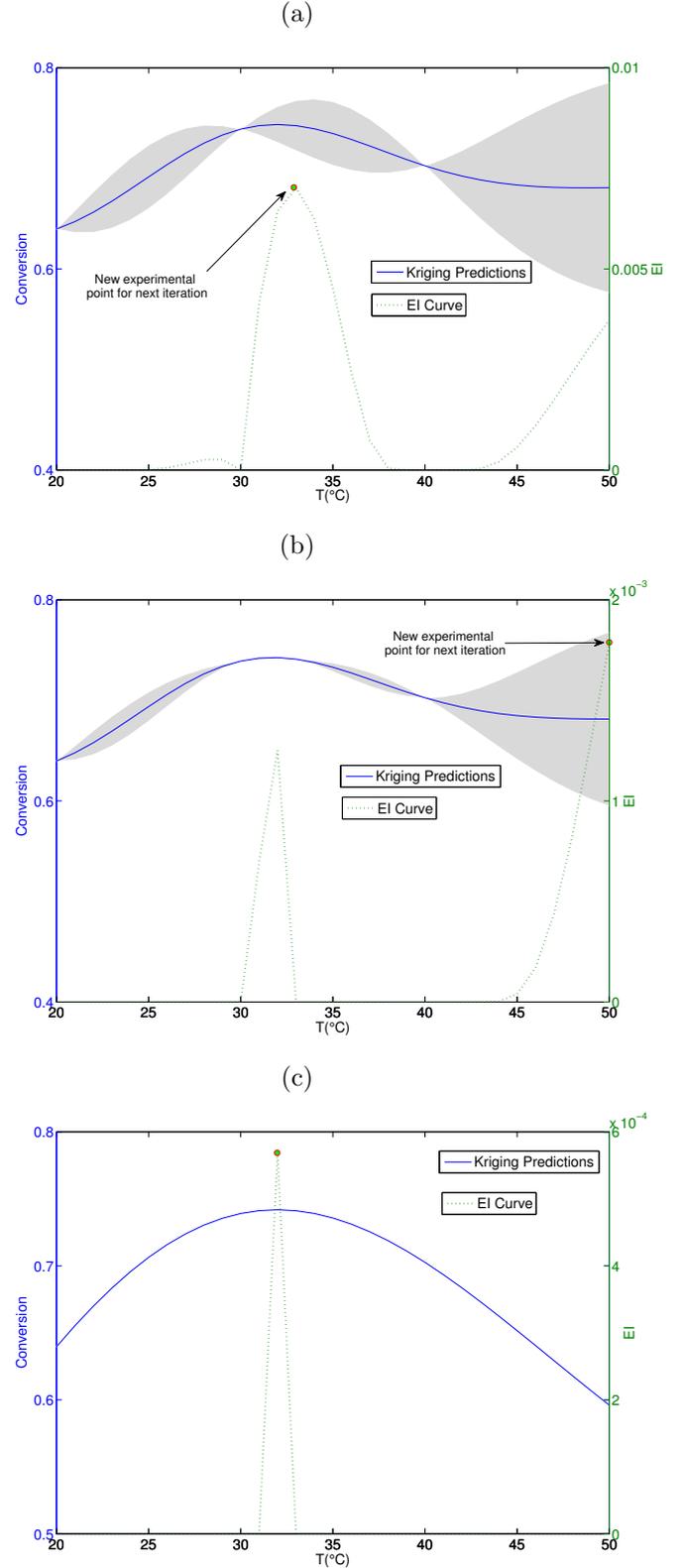
Fig. 2. Kriging model prediction and its 95% confidence interval (shaded area) at the first three iterations. The expected improvement (EI) used for allocating new experiments is also illustrated.

is assigned to 50 °C at the second iteration. Up to this point, a total of five experiments have been tried, and the prediction as given in Fig. 2(c) has small uncertainty

464
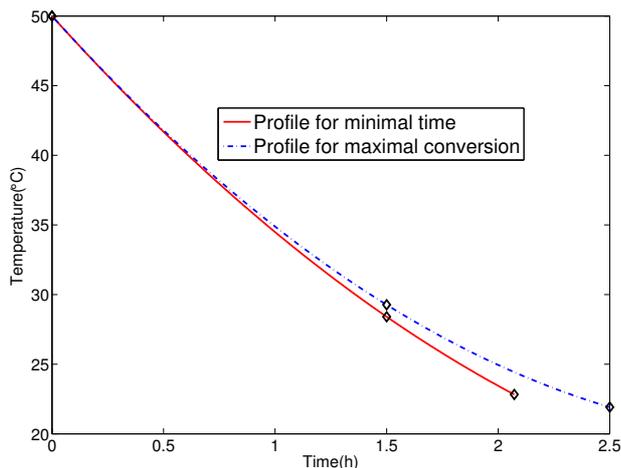
Fig. 3. Optimal temperature profiles.

Table 2. Maximal conversion obtained through optimization.

Method	No. Experiments	Conversion
Isothermal	7	74.2%
Dynamic	12	77.5%

left. In real practice, the experimenter may choose to terminate the algorithm here. To gain more confidence, we have selected a relatively small threshold for EI, and thus the algorithm proceeded with two more iterations. The optimal temperature is identified to be 32.0 °C, corresponding to a conversion of 74.2%.

Next we consider a time-varying temperature profile by using the orthogonal collocation method. Three collocation points are allocated at three time steps: 0, 1.5 and 2.5 h. Hence the process factors to be desired are three temperature values, $[T(0), T(1.5), T(2.5)]$, at these time steps. Initially HSS is used to give nine design points. Then by following the proposed method, only three more iterations are needed and the identified optimal temperature values at the three collocation points are $[50, 29.3, 21.9]$, corresponding to a conversion rate of 77.5%. The optimal temperature profile is illustrated in Fig. 3 as dash-dotted line. The optimal profile is a decreasing one, which is consistent with the theoretical analysis in (Georgakis, 2009). Table 2 compares the maximal conversion obtained through optimization. At the expense of several more experiments, the dynamic temperature profile can achieve higher conversion of the reactant.

3.2 Minimize the reaction time at a given conversion

The other common goal in batch process optimization is to minimize the reaction time at a given conversion, which is usually set to be less than the maximally achievable value. This strategy is well justified if the benefit from reduced processing time exceeds the loss in yield. In this work, the objective is to obtain the minimal reaction time when conversion reaches 75%. To account for the possibility that some poor choice of temperature would never reach the conversion rate of 75%, we artificially set the longest reaction time to be 5.5 h. This is in analogous to the real practice, whereby if the conversion rate does not achieve

Table 3. Minimal batch duration obtained through optimization.

Method	No. Experiments	Duration (h)
Isothermal	7	2.65
Dynamic	16	2.08

.
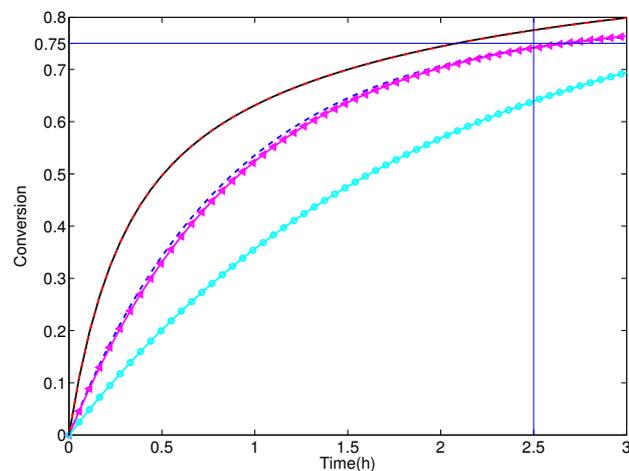
Fig. 4. The trajectory of conversion at different operating conditions. The two profiles of dynamic operations are almost non-discernible.

a desired value for a long time, then the experiment will simply be terminated.

Similar to the previous case, we first consider isothermal operation, using the same three design points for temperature between 20 and 50 °C for developing the initial kriging model. The iterative optimization strategy terminates after four iterations, resulting in an optimal temperature of 30.7 °C with reaction time 2.65 h.

The time-varying temperature profile is implemented by allocating three orthogonal points at 0, 1.5 and 5.5 h (the last corresponding to the specified longest reaction time). Similar to maximizing conversion, initially nine data points were obtained. At the convergence, a total of 16 runs of the simulation were needed to obtain the minimal time of 2.08 h, which is remarkably better than isothermal operation (see Table 3).

The optimal temperature profile is also given in Fig. 3. The two temperature profiles, corresponding to different optimization objectives, are initially very similar and slightly different after about 1 h. Fig. 4 illustrates the conversion profiles for isothermal and dynamic operations. A "baseline" profile, corresponding to using a fixed temperature of 20 °C (the lower bound of the range), is also depicted. Again, this figure confirmed the better performance of dynamic operations. More interestingly, for both isothermal and dynamic cases, it appears that maximizing conversion and minimizing reaction time gave almost the same conversion profiles. In other words, should we conduct multi-objective optimization by considering the two objectives simultaneously, we would achieve largely similar results by two single-objective optimizations. Certainly, this conclusion cannot be directly generalized to other processes. Data-based modelling and multi-objective optimization,

where the model uncertainty is considered, is currently under investigation.

4. CONCLUSION

In this work, we have demonstrated an iterative approach to data-based process modelling and optimization. We adopted the kriging model (Gaussian process) that is capable of providing both accurate mean prediction and reliable prediction variance. For data-based modelling, the prediction variance can be effectively reduced by conducting more experiments. We suggested to use a combined criterion, the expected improvement, in an attempt to allocate new experiments to under-explored region (with high prediction variance) as well as to search the best process performance. We further extended the method to handle time-varying process factors. The case study has indicated the effectiveness of the proposed method. In principle, the proposed method is equally applicable to general processes, where a response variable is affected by several factors, be they static or dynamic.

Currently, we are working on the application of this method to the design of catalysts for several energy-related processes, including dry reforming of methane and Fischer-Tropsch synthesis.

REFERENCES

Apley, D., Liu, J., and Chen, W. (2006). Understanding the effects of model uncertainty in robust design with computer experiments. *Journal of Mechanical Design*, 128, 945–958.

Dutta, J.R., Dutta, P.K., and Banerjee, R. (2004). Optimization of culture parameters for extracellular protease production from a newly isolated pseudomonas sp. using response surface and artificial neural network models. *Process Biochemistry*, 39, 2193–2198.

Fang, K.T., Lin, D.K.J., Winker, P., and Zhang, Y. (2000). Uniform design: theory and application. *Technometrics*, 42, 237–248.

Georgakis, C. (2009). A model-free methodology for the optimization of batch processes: Design of dynamic experiments. In *Proc. 8th International Symposium on Dynamics and Control of Process Systems (DYCOPS)*. Istanbul, Turkey.

Hadjmohammadi, M. and Kamel, K. (2008). Response surface methodology and support vector machine for the optimization of separation in linear gradient elution. *Journal of Separation Science*, 31, 3864–3870.

Jones, D. (2001). A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization*, 21(4), 345–383.

Kalagnanam, J. and Diwekar, U. (1997). An efficient sampling technique for off-line quality control. *Technometrics*, 39, 308–319.

Klatt, K. and Marquardt, W. (2009). Perspectives for process systems engineeringPersonal views from academia and industry. *Computers and Chemical Engineering*, 33(3), 536–550.

McKay, M.D., Beckman, B.J., and Conover, W.J. (1979). A comparison of three methods for selecting values for input variables in the analysis of output from a computer code. *Technometrics*, 21, 239–245.

Myers, R.H. and Montgomery, D.C. (1995). *Response Surface Methodology*. Wiley.

Sacks, J., Welch, W., Mitchell, T., and Wynn, H. (1989). Design and analysis of computer experiments. *Statistical Science*, 4, 409–423.

Shao, P., Jiang, S.T., and Ying, Y.J. (2007). Optimization of molecular distillation for recovery of tocopherol from rapeseed oil deodorizer distillate using response surface and artificial neural network models. *Food and Bioproducts Processing*, 85, 85–92.

Tang, Q., Lau, Y., Hu, S., Yan, W., Yang, Y., and Chen, T. (2010). Response surface methodology using Gaussian processes: towards optimizing the trans-stilbene epoxidation over Co^{2+}-NaX catalysts. *Chemical Engineering Journal*, 156, 423–431.

Villadsen, J. and Michelsen, M. (1978). *Solution of differential equation models by polynomial approximation*. Prentice Hall.

Yuan, J., Wang, K., Yu, T., and Fang, M. (2008). Reliable multi-objective optimization of high-speed WEDM process based on Gaussian process regression. *International Journal of Machine Tools and Manufacture*, 48, 47–60.