

Bayesian migration of Gaussian process regression for rapid process modeling and optimization

Wenjin Yan^a, Shuangquan Hu^a, Yanhui Yang^a, Furong Gao^b, Tao Chen^{*,a}

^a*School of Chemical and Biomedical Engineering, Nanyang Technological University, 62 Nanyang Drive, Singapore 637459, Singapore*

^b*Department of Chemical and Biomolecular Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong*

Abstract

Data-based empirical models, though widely used in process optimization, are restricted to a specific process being modeled. Model migration has been proved to be an effective technique to adapt a base model from a old process to a new but similar process. This paper proposes to apply the flexible Gaussian process regression (GPR) for empirical modeling, and develops a Bayesian method for migrating the GPR model. The migration is conducted by a *functional* scale-bias correction of the base model, as opposed to the restrictive *parametric* scale-bias approach. Furthermore, an iterative approach that jointly accomplishes model migration and process optimization is presented. This is in contrast to the conventional “two-step” method whereby an accurate model is developed prior to model-based optimization. A rigorous statistical measure, the expected improvement, is adopted for optimization in the presence of prediction uncertainty. The proposed methodology has been applied to the optimization of a simulated chemical process, and a real catalytic reaction for the epoxidation of *trans*-stilbene.

Key words:

Design of experiments, Gaussian process regression, heterogeneous catalysis, model adaptation, model uncertainty, response surface methodology

1. Introduction

In the discipline of process systems engineering, mathematical models form the cornerstone for the design, optimization and control of chemical processes [1]. In general, process models represent the relationship between input variables \mathbf{x} (also termed factors) and responses (also termed outputs) \mathbf{y} , and they may be classified into three categories: first-principles, empirical and hybrid (a combination of first-principles and empirical models). First-principles models are based on fundamental chemical/physical mechanisms, and thus they provide phenomenological insight of the process and attain good extrapolation capability. However, developing such a model requires considerable experience and computation. Alternatively, empirical models are developed purely based on experimental data and are typically valid to a specific process and reliable only within the operating region where the data are collected. Nevertheless, empirical models are advantageous in terms of the simplicity in model development and implementation. They are especially applicable to early-stage process development when first-principles modeling is usually not cost-effective. The scope of this paper is restricted to such scenarios to develop process models for off-line (as opposed to on-line) optimization of a certain process. This technique is also termed *response surface methodology* (RSM), which is a collection of proper design of experiments (DoE), empirical modeling, and model-based optimization methods [2]. In the literature, the choice of empirical models has included artificial neural network (ANN)

*Corresponding author. Tel.: +65 6513 8267; Fax: +65 6794 7553.
Email address: chentao@ntu.edu.sg (Tao Chen)

[3], least-squares support vector machines (LS-SVMs) [4] and Gaussian process regression (GPR, also known as kriging model) [5, 6, 7, 8, 9, 10].

A major challenge for empirical modeling is the lack of extrapolating capability. Extrapolation refers to both the change of factors’ range that has not been explored, and the change of the process itself. The latter case is of primary interest in this study. If the *new* process resembles the *base* (old) process in a certain way, then it is reasonable to expect that the base and new models behave similarly. According to [11], a process may be represented by either descriptive attributes or quantitative models, upon which similarity was defined and discussed in detail. Similar processes may arise when different oxidants are tested on the same catalytic reaction, resulting in different reaction mechanisms and/or rates (see the example in Section 3.2). Another example was described in [12] to make similar polymer products with different shapes and types through injection molding. To fully exploit the similarity between processes for modeling, a strategy was proposed in [12] to migrate the base model to the new process such that one of the following outcomes is expected:

- (i) To attain similar prediction accuracy, fewer data are required for migrating to the new model than for developing the base model; Or,
- (ii) the migrated new model is more accurate than the base model if using (nearly) equal number of experimental data for both processes.

Usually, the former situation is desired to save time and cost.

The central component of the migration technique is to formulate the new model in terms of the prediction from the base model. In particular, the following method was adopted in [12]:

$$\mathbf{y}_{\text{new}} = s_{\text{O}}f(s_{\text{I}}\mathbf{x}_{\text{new}} + b_{\text{I}}) + b_{\text{O}} \quad (1)$$

where $f(\cdot)$ denotes the base model, s and b are scale and bias parameters, respectively, and the subscripts I and O are short for input and output, correspondingly. In addition, \mathbf{x}_{new} and \mathbf{y}_{new} refer to data from the new process. The scale and bias parameters may be estimated by minimizing the squared prediction errors. More recently, other migration strategies have been explored, including the local modeling approach [13] and ensemble modeling for dynamic models [14, 11]. In these previous studies, either quadratic polynomials [12, 13] or ANNs [14, 11] were chosen for empirical modeling.

In this paper, we suggest to use GPR in place of polynomials or ANNs for process modeling and model migration, since GPR has been shown to attain more reliable predictive performance when compared with other popular modeling methods [15, 16, 10, 9]. In addition, we propose to replace the output scale-bias *parameters* in Eq. (1) (s_{O} and b_{O}) by scale-bias *functions* of process factors. This is motivated by the fact that the parametric scale-bias updating rule may not possess sufficient flexibility to model the process of interest. As a simple and hypothetical example, suppose that the base model $f(\mathbf{x})$ is a linear function of \mathbf{x} . Then, parametric scale-bias updating will only result in a linear model, which does not offer the flexibility to account for new processes that are non-linear. As a result, the updated model may not converge to the true response surface asymptotically (i.e. when the number of experimental data goes to infinity). In the literature, the method of scale-bias functions was adopted for integrating high-accuracy (but computationally slow) and low-accuracy (but fast) computer simulations [17]. In the present study, a Bayesian migration approach is developed to update the scale-bias functions, given experimental data from the new process. The Bayesian methodology is desired to fully incorporate the uncertainties in the model parameters. The case studies in this paper suggest that when compared with a model purely developed from new process data, such a migration strategy (i) is superior when experimental data are limited; and (ii) gives similar accuracy when abundant data are available.

In addition, the current model migration studies [11, 12, 13, 14] are primarily focused on improving the prediction accuracy of the new model across the entire range of the factors. If the objective of modeling is to identify the process factors that give the optimal process response, then the new model will be subsequently used for optimization purpose. This two-step method is conceptually non-ideal,

since improving prediction accuracy at the region of low response value does not directly contribute to optimizing the process. Throughout this paper, the objective is assumed to maximizing a certain response value. A one-step iterative method within which model migration and process optimization are jointly accomplished is required, for rapid optimization purpose which can save time and resource.

The other major contribution of this paper is to achieve model migration and process optimization simultaneously. A straightforward method may be to find a process optimum based on the prediction of the migrated model, and then conduct new experiment at this optimum. Subsequently, the model is further migrated by including all available data from the new process. The procedure continues until the improvement of process response becomes smaller than a threshold. However, this approach ignores the fact that model prediction is not perfect and carries uncertainty, which is represented by prediction variance of the model. Note that the Bayes origin of GPR automatically provides the uncertainty (variance) of prediction [7]. A large variance suggests that the experimental data around this point are not sufficient to give a reliable prediction. Therefore, both predictive mean and variance should be jointly considered; otherwise, the algorithm is likely to find only a local, not global, optimum [5]. In the literature, a few methods have been suggested to handle prediction uncertainty, including maximization of lower or upper prediction bound [9, 10], minimization of information free energy [18, 19], maximization of relative information gain [20], and maximization of expected improvement (EI) [5]. The criterion of EI incorporates the predictive mean and variance in a rigorous statistical framework, as opposed to user-determined weighting of the impact of mean and variance in other methods. Therefore, EI is adopted in this study to identify the factors’ values that are likely to improve the process response and/or prediction accuracy.

The remainder of this paper is organized as follows. Section 2 presents the overall framework of the joint migration-optimization method, including a brief introduction to GPR, the selection of initial design points through clustering technique for experimenting the new process, the migration strategy and the EI-based optimization method. Section 3 validates the proposed method using a simulated chemical process and an experimental catalytic reaction for the epoxidation of *trans*-stilbene. Finally, Section 4 concludes this paper.

2. Model migration and process optimization

The overall approach of joint model migration and process optimization is illustrated in Figure 1. Suppose that the old process has been experimented and a certain amount of data have been collected. Ideally, the old process should be investigated systematically where the values of process factors are determined according to rigorous design of experiments (DoE) techniques. The classical fractional factorial and central composite designs are effective when combined with polynomial regression models [2]; however, they usually perform unsatisfactorily when being used to develop more complex models (such as ANN and GPR) due to inadequate coverage of the factors’ space [21]. The recognition of this disadvantage of classical DoEs has motivated the concept of “space-filling” designs that allocate design points to be uniformly distributed within the range of each factor.

(Figure 1 about here)

The most widely used space-filling designs include Latin hypercube sampling (LHS) [22], uniform design (UD) [23] and Hammersley sequence sampling (HSS) [24]. It was shown that UD and HSS provide more uniform coverage of the design space than LHS [23, 24]. In addition, HSS is more straightforward to implement than the optimization-based UD, while the two methods achieve similar performance [15]. Therefore, HSS is adopted in this study for experimental design, and its implementation details are given in Section 2.1.

Once the experimental data are collected according to DoE (Section 2.1), a base GPR model can be developed for the old process. In Section 2.2, a brief overview of GPR will be given.

When a new yet similar process needs to be investigated, the base model is to be migrated for the new process. The proposed migration strategy is as follows, and the various components will be discussed subsequently.

1. Apply clustering technique to the data from the old process to determine the initial experimental points for the new process (Section 2.3).
2. Migrate the base GPR model using all available data from the new process (Section 2.4).
3. Allocate a new experiment by maximizing the EI that is computed from the migrated model, and then conduct the experiment accordingly (Section 2.5).
4. Terminate when process improvement is small or further experimentation is prohibited due to limited resources; otherwise go to step 2.

2.1. Hammersley sequence sampling (HSS)

The basis of HSS design is that any integer n can be written in a radix notation of another integer R as follows:

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

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

$$\begin{aligned} \psi_R(n) &= 0.n_m n_{m-1} \cdots n_2 n_1 n_0 \\ &= n_m R^{-1} + n_{m-1} R^{-2} + \cdots + n_1 R^{-m} + n_0 R^{-m-1} \end{aligned} \quad (3)$$

Therefore, if n design points, each being a vector of dimension d , need to be allocated by HSS design, the first $d-1$ prime numbers should be selected as the integer R in Eq. (2): $R_1, R_2, \cdots, R_{d-1}$. The design points are given by

$$\mathbf{x}_i = \mathbf{1} - \left[\frac{i}{n}, \psi_{R_1}(n), \psi_{R_2}(n), \cdots, \psi_{R_{d-1}}(n) \right]^T \quad (4)$$

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

2.2. Developing a GPR model

GPR, originally initiated in the statistical community [6], has gained significant attention as a powerful modeling tool for general scientific and engineering tasks [9, 10, 16, 25, 26, 27, 28]. When compared with other popular modeling methods like ANN, GPR has been demonstrated to give more reliable predictive performance [10, 7]. In addition, the Bayes origin of GPR automatically provides the uncertainty (variance) of prediction, which is indispensable to robust model-aided optimization [9, 10, 5]. Specifically, GPR aims at relating the scalar process response y to the d -dimensional factor \mathbf{x} . If multiple responses are present, a common practice is to develop separate GPR models for each response, while other approaches may also be used to address this issue [29, 30]. Given a set of experimental data of size n : $\{\mathbf{x}_i, y_i; i = 1, \dots, n\}$, a GPR is defined such that the regression function $y(\mathbf{x})$ has a Gaussian prior distribution with zero mean, or in discrete form:

$$\mathbf{y} = (y_1, \dots, y_n)^T \sim G(\mathbf{0}, \mathbf{C}) \quad (5)$$

where \mathbf{C} is an $n \times n$ covariance matrix, whose ij -th element is defined by a covariance function: $\mathbf{C}_{ij} = C(\mathbf{x}_i, \mathbf{x}_j)$. A widely used covariance function is

$$C(\mathbf{x}_i, \mathbf{x}_j) = a_0 + a_1 \sum_{k=1}^d x_{ik} x_{jk} + v_0 \exp \left(- \sum_{k=1}^d w_k (x_{ik} - x_{jk})^2 \right) + \sigma^2 \delta_{ij} \quad (6)$$

where x_{ik} is the k -th variable of \mathbf{x}_i ; $\delta_{ij} = 1$ if $i = j$, otherwise $\delta_{ij} = 0$. The four terms in Eq. (6) account for the effect of constant bias, linear correlation, non-linear correlation and random noise, respectively. Such a covariance function was shown to attain excellent prediction performance when compared with other forms [7]. The covariance function is parameterized by $\boldsymbol{\theta} = (a_0, a_1, v_0, w_1, \dots, w_d, \sigma^2)^T$, which is termed ‘‘hyper-parameters’’ due to the origin in Bayesian non-parametric statistics. As such, GPR may also be derived from the principle of Bayesian non-parametric regression.

For a new data point with factor \mathbf{x}^* , the predicted response y also follows a normal distribution, of which the mean (\hat{y}^*) and variance ($\sigma_{\hat{y}^*}^2$) are

$$\hat{y}^* = \mathbf{k}^T(\mathbf{x}^*)\mathbf{C}^{-1}\mathbf{y} \quad (7)$$

$$\sigma_{\hat{y}^*}^2 = C(\mathbf{x}^*, \mathbf{x}^*) - \mathbf{k}^T(\mathbf{x}^*)\mathbf{C}^{-1}\mathbf{k}(\mathbf{x}^*) \quad (8)$$

where $\mathbf{k}(\mathbf{x}^*) = [C(\mathbf{x}^*, \mathbf{x}_1), \dots, C(\mathbf{x}^*, \mathbf{x}_n)]^T$.

The hyper-parameters can be estimated by maximizing the logarithm of the likelihood function defined in Eq. (5): $L = \log p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{X})$. To solve this non-linear optimization problem, the derivative of log-likelihood with respect to each hyper-parameter θ is usually needed and calculated as follows:

$$\frac{\partial L}{\partial \theta} = -\frac{1}{2} \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta} \right) + \frac{1}{2} \mathbf{y}^T \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta} \mathbf{C}^{-1} \mathbf{y} \quad (9)$$

where $\partial \mathbf{C} / \partial \theta$ can be calculated from the covariance function. A Matlab implementation of GPR, based on the conjugate gradient method, is publicly available from <http://www.gaussianprocess.org/gpml/code/matlab/doc/> [7], and it is used in this study. In addition, to ensure numerical stability, the data should be normalized to the range of $[-1, 1]$ at each process factor.

2.3. Clustering to allocate initial experiments for the new process

A straightforward approach to allocating initial experiments for the new process is to apply a space-filling DoE method, similar to the design of the old process, with the aim to obtain a fair coverage of the factors’ space [12, 14, 11]. Later, this approach was recognized to ignore important information that could lead to better initial design [13]. In particular, space-filling DoE only assures a uniform coverage of the factors’ space; a better approach is to consider both the factors’ and response’s spaces. For example, consider an illustrative example where the response y is a function of a single factor x : $y = \exp(-x) \times \sin(3x) + 0.3$. Figure 2(a) shows the HSS-allocated six design points that are uniformly distributed within the interval $x \in [0, 5]$. By jointly considering x and y , the design can be improved by allocating more points at the region where y changes rapidly with x , and assigning fewer points where y is a relatively smooth function of x (Figure 2(b)).

(Figure 2 about here)

Nevertheless, the response’s values are only available *after* the factors have been designed and experiments have been conducted. Fortunately, in the context of model migration, the old process has provided the data with both \mathbf{x} and \mathbf{y} . Under the assumption that the old and new processes are similar, a reasonable method is to select a subset of the old data that have a uniform coverage of both \mathbf{x} and \mathbf{y} space, and then use these \mathbf{x} values as initial design for the new process. A rigorous method to assess process similarity was presented in [11] and can be used in practice to test this assumption. In addition, the initial design method needs to be incremental, since if the initial experimental data do not give a satisfactorily migrated model, then more design points will be needed. To fulfill these requirements, an iterative clustering algorithm, originally proposed by Chiu [31] and recently used in [13], is adopted in this study. Clustering techniques group the data into distinct clusters based on the distance at both \mathbf{x} and \mathbf{y} spaces, and the resultant cluster centers become good candidates for the initial experiments. The original clustering algorithm [31] uses the concatenated vector, $\mathbf{z} = [\mathbf{x}; \mathbf{y}]$,

which may under-weight the impact of response variable if multiple factors are present. This issue is mitigated by considering the dimensions of \mathbf{x} and \mathbf{y} when calculating the distance between data points (details are given in Eq. (10)). The adopted algorithm is incremental by selecting one data point as the cluster center at a time, and adding subsequent cluster centers does not affect the previously obtained centers. This is in contrast to the classic k -means and other clustering techniques, which require to modify the previously selected cluster centers when adding more clusters [32].

Next, a brief overview of the clustering method is presented. Consider a dataset $\{(\mathbf{x}_i, \mathbf{y}_i), i = 1, \dots, n\}$ where the factor \mathbf{x}_i is of d_1 dimensions and the respective response \mathbf{y}_i is of d_2 dimensions. The clustering algorithm initiates by calculating the *potential* of each data point being a cluster center. An effective measure of this potential, adapted from [31] by considering the dimensions of both \mathbf{x} and \mathbf{y} , is given by

$$P_i = \sum_{j=1}^n \exp \left(-4 \times \left(\frac{1}{d_1} \|\mathbf{x}_i - \mathbf{x}_j\|^2 + \frac{1}{d_2} \|\mathbf{y}_i - \mathbf{y}_j\|^2 \right) \right) \quad (10)$$

Essentially, the “potential function” P_i measures the distance of the i -th data point to all other data. The data point with the highest potential is chosen as the first cluster center, denoted by $(\mathbf{x}_1^*, \mathbf{y}_1^*)$ with the corresponding potential value of P_1^* . The potential of the rest data is then updated as

$$P_i \Leftarrow P_i - P_1^* \exp \left(-\frac{4}{\beta^2} \times \left(\frac{1}{d_1} \|\mathbf{x}_i - \mathbf{x}_1^*\|^2 + \frac{1}{d_2} \|\mathbf{y}_i - \mathbf{y}_1^*\|^2 \right) \right) \quad (11)$$

where β is a tuning parameter for adjusting the potential P_i according to its distance to the first cluster center. To avoid collecting the data near the former cluster centers, β should be a constant larger than 1. Chiu suggested a value of 1.5 [31], which is used in this study.

After updating all the potentials according to Eq. (11), the data point with the highest potential is selected as the second cluster center. By iterating this updating-selection procedure, a total of n_{initial} cluster centers are selected as the initial design for experimenting the new process.

The number of initial experiments, n_{initial} , is a rather subjective choice. In practice, it is usually determined by available experimental resources and the quality of the migrated model. If the model does not provide satisfactory prediction, which may be judged by leave-one-out cross-validation, then more cluster centers (thus more design points for the new process) are needed. The incremental nature of the initial experimental design clearly requires an incremental clustering method, e.g. the one presented in this study, as opposed to the classical k -means algorithm.

2.4. Model migration

Given the initial design for the new process, a set of experimental data are collected and denoted as $\{(\mathbf{x}_i, y_i)_{\text{new}}, i = 1, \dots, m\}$. These data form the basis for developing a migrated model for the new process. Since this and the next sub-sections primarily discuss the migration and optimization of the new process, the subscript “new” is neglected unless confusions may arise. Suppose that a base GPR model, developed for the old process, predicts at a data point \mathbf{x}_i of the new process to be $z(\mathbf{x}_i) \sim G(z_i, s_i^2)$, whereby the predictive mean and variance are given by Eqs. (7)(8). The objective of model migration is to formulate a new model on top of the base model for the new process. Specifically, the new model is structured with scale-bias correcting functions as follows [17]:

$$y(\mathbf{x}_i) = \rho(\mathbf{x}_i)z_i + \delta(\mathbf{x}_i), \quad i = 1, \dots, m \quad (12)$$

The bias adjustment $\delta(\mathbf{x}_i)$ is chosen as a zero-mean Gaussian process with covariance function parameterized by “hyper-parameters” $\boldsymbol{\theta}$ as in Eq. (6). The scale correction is chosen to be a linear function:

$$\rho(\mathbf{x}_i) = \rho_0 + \sum_{j=1}^d \rho_j x_{ij} \quad (13)$$

where d is the dimension of process factors \mathbf{x} and accordingly x_{ij} is the j -th element of \mathbf{x}_i . Since $\delta(\cdot) \sim G(\mathbf{0}, \mathbf{C})$ is a GPR, the new model is also a GPR with discrete form as:

$$\mathbf{y} = [y_1, \dots, y_m]^T \sim G(\mathbf{F}\boldsymbol{\rho}, \mathbf{C}) \quad (14)$$

where

$$\mathbf{F} = \begin{pmatrix} z_1 & x_{11}z_1 & x_{12}z_1 & \cdots & x_{1d}z_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_m & x_{m1}z_m & x_{m2}z_m & \cdots & x_{md}z_m \end{pmatrix}, \quad \boldsymbol{\rho} = \begin{pmatrix} \rho_0 \\ \vdots \\ \rho_d \end{pmatrix} \quad (15)$$

Previous study showed that a simple linear scale function, in conjunction with a GPR bias function $\delta(\cdot)$, is sufficiently flexible for modeling purpose, since the ultimate model in Eq. (14) is non-parametric [17]. In addition, Kennedy and O'Hagan [33] indicated that more complex form of the scale function is usually not necessary under a broader context of GPR.

In [17], a point estimation was adopted to obtain $\boldsymbol{\rho}$, and the resultant predictive distribution does not account for the uncertainty of the estimated $\boldsymbol{\rho}$. In this study, a Bayesian approach is proposed to integrate out the regression coefficients $\boldsymbol{\rho}$, which is an effective method to fully incorporate the parameter uncertainty [33]. In particular, independent prior distribution is assigned for each element of $\boldsymbol{\rho}$: $\rho_j \sim G(0, \alpha^2)$, then

$$p(\mathbf{y}|\alpha, \boldsymbol{\theta}) = \int p(\mathbf{y}|\boldsymbol{\rho}, \boldsymbol{\theta})p(\boldsymbol{\rho}|\alpha)d\boldsymbol{\rho} = G(\mathbf{0}, \alpha^2\mathbf{F}\mathbf{F}^T + \mathbf{C}) \quad (16)$$

Therefore, as opposed to estimate the regression coefficients $\boldsymbol{\rho}$ directly, we estimate α^2 and the hyper-parameters $\boldsymbol{\theta}$ in \mathbf{C} . They can be obtained by maximizing the logarithm of the likelihood function in Eq. (16): $\log p(\mathbf{y}|\alpha, \boldsymbol{\theta})$. Similar to the base GPR model development, this is also a non-linear optimization problem that can be solved by using gradient-based methods. The derivatives of log-likelihood with respect to each element of $\boldsymbol{\theta}$ and α^2 are readily available by replacing \mathbf{C} with $\alpha^2\mathbf{F}\mathbf{F}^T + \mathbf{C}$ in Eq. (9) and noting $\partial(\alpha^2\mathbf{F}\mathbf{F}^T + \mathbf{C})/\partial(\alpha^2) = \mathbf{F}\mathbf{F}^T$. Finally, the predictive mean (\hat{y}^*) and variance ($\sigma_{\hat{y}^*}^2$) at a new data point \mathbf{x}^* are also given in Eqs. (7)(8) with \mathbf{C} being replaced by $\alpha^2\mathbf{F}\mathbf{F}^T + \mathbf{C}$.

2.5. Iterative optimization

As discussed in the Introduction, the main objective of this work is to develop a one-step method to jointly achieve model migration and process optimization. A simple method is to find an optimum \mathbf{x} based on the prediction from the migrated model, and then conduct new experiment at \mathbf{x} which will be used to further migrate the process model. However, this approach ignores the unavoidable model-reality mismatch, which has been quantified by the GPR model in terms of predictive variance. For example, a data point with an inferior mean prediction yet large predictive variance may ultimately give a favorable response in reality. A large predictive variance generally suggests that the experiment data around this point are not sufficient for the model to give a reliable prediction. Therefore, both predictive mean and variance need to be considered in the optimization algorithm. In the present study, the statistical measure of expected improvement (EI) [5] is adopted to allocate subsequent experiments at the region where the predictive mean is high and/or the predictive variance is large.

As the name suggests, EI quantifies the improvement that is expected to obtain by conducting experiment at a design point. Specifically, let $y(\mathbf{x})$ be the prediction at \mathbf{x} from a GPR model. For a maximization problem, the predicted improvement over the best response obtained through experiments so far (denoted f_{best}) is thus $I(\mathbf{x}) = y(\mathbf{x}) - f_{\text{best}}$. Since $y(\mathbf{x})$ is Gaussian distributed with mean \hat{y} and variance $\sigma_{\hat{y}}^2$, the improvement $I(\mathbf{x})$ is also Gaussian distributed with mean $\hat{y} - f_{\text{best}}$ and the same variance. Note that the mean (or expectation) of I is defined as $\int_{-\infty}^{\infty} Ip(I)dI$ and is simply $\hat{y} - f_{\text{best}}$. However, improvement requires a positive I , i.e. $y(\mathbf{x})$ should be greater than the best response so far (f_{best}). Hence, the expectation is calculated by integration from 0 to ∞ as follows:

$$\text{EI}(\mathbf{x}) = E[\max\{0, I(\mathbf{x}^*)\}] = \int_0^\infty Ip(I)dI \quad (17)$$

Substituting the probability density function of Gaussian distribution into above equation gives

$$\text{EI}(\mathbf{x}) = \int_0^\infty I \left\{ \frac{1}{\sqrt{2\pi}\sigma_{\hat{y}}} \exp \left[-\frac{(I - \hat{y} + f_{\text{best}})^2}{2\sigma_{\hat{y}}^2} \right] \right\} dI \quad (18)$$

After integration by parts, the expectation of improvement at \mathbf{x} is given by:

$$\text{EI}(\mathbf{x}) = \sigma_{\hat{y}}[u\Phi(u) + \phi(u)] \quad (19)$$

where $u = (\hat{y} - f_{\text{best}})/\sigma_{\hat{y}}$, $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution function and density function of the standard normal distribution, respectively:

$$\Phi(u) = \frac{1}{2} \text{erf} \left(\frac{u}{\sqrt{2}} \right) + \frac{1}{2} \quad (20)$$

$$\phi(u) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{u^2}{2} \right) \quad (21)$$

EI will increase if the mean prediction is greater than f_{best} and/or the predictive variance is large, and thus further experiments should be conducted at this region. Therefore, we search for process factors that maximize EI. For this purpose, the derivative of EI with respect to \mathbf{x} is needed and calculated as [34]:

$$\frac{\partial \text{EI}(\mathbf{x})}{\partial \mathbf{x}} = [u\Phi(u) + \phi(u)] \frac{\partial \sigma_{\hat{y}}(\mathbf{x})}{\partial \mathbf{x}} + \sigma_{\hat{y}}(\mathbf{x})\Phi(u) \frac{\partial u}{\partial \mathbf{x}} \quad (22)$$

where

$$\frac{\partial \sigma_{\hat{y}}(\mathbf{x})}{\partial \mathbf{x}} = - \left(\frac{\partial \mathbf{k}^T}{\partial \mathbf{x}} \mathbf{C}^{-1} \mathbf{k} \right) / \sigma_{\hat{y}}(\mathbf{x}) \quad (23)$$

$$\frac{\partial u}{\partial \mathbf{x}} = \left(\frac{\partial \mathbf{k}^T}{\partial \mathbf{x}} \mathbf{C}^{-1} \mathbf{y} - u \frac{\partial \sigma_{\hat{y}}(\mathbf{x})}{\partial \mathbf{x}} \right) / \sigma_{\hat{y}}(\mathbf{x}) \quad (24)$$

The Jacobian $\partial \mathbf{k}^T / \partial \mathbf{x}$ depends on the form of the covariance function. Alternatively, the method of finite difference may be used to calculate $\partial \text{EI}(\mathbf{x}) / \partial \mathbf{x}$, and it may be more straightforward to implement in practice.

3. Case study

3.1. A simulated chemical process

A benchmark [35, 24] of a continuously stirred tank reactor (CSTR) is used in simulation to illustrate and validate the proposed algorithm. The simulation describes a chemical reaction $A \rightarrow B \rightarrow C$, where B is the desired product. The mechanistic model assumes steady-state operation and is defined by the following equations:

$$\tau = \frac{-k_A k_B C_{Bi} + \sqrt{[k_A^2 k_B^2 C_{Bi}^2 + k_A^2 k_B (C_{Ai} + C_{Bi}) \times (k_A C_{Ai} - k_B C_{Bi})]}}{k_A^2 k_B (C_{Ai} + C_{Bi})}$$

$$C_A = \frac{C_{Ai}}{1 + k_A \tau}$$

$$C_B = \frac{C_{Bi} + k_A C_A \tau}{1 + k_B \tau}$$

$$F = V / \tau$$

$$rate = C_B F / 1000$$

Following previous studies on this process [24, 35], four process factors are considered: inlet concentration of A (C_{Ai}), inlet concentration of B (C_{Bi}), reactor temperature (T) and reactor volume (V). Production rate of B ($rate$) is the target response which needs to be maximized. C_A and C_B are concentrations of A and B in the outlet stream, respectively. F denotes the volumetric flow through the CSTR. Two similar processes are simulated by varying the reaction kinetic parameters, k_A and k_B , as shown in Table 1. The range of the process factors are listed in Table 2. It should be noted that the range of C_{Ai} is shifted and the range of T is expanded for the new process, to demonstrate model migration under extrapolation.

(Table 1 and Table 2 about here)

The base GPR model is developed by simulating the CSTR at 100 design points assigned according to HSS. A random noise of 1% standard deviation is added to the production rate to simulate measurement errors. To validate the model migration algorithm, the design points for the new process are first allocated by HSS; the use of clustering will be demonstrated later for optimization purpose. Different amount of new design points for model migration are tested. The prediction capability of the model is evaluated on additional 200 data, also designed by HSS, using the criterion of root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (25)$$

where n is the number of data points; y_i is the actual response (from either process simulation or experiments) and \hat{y}_i is the prediction. RMSE may be interpreted as the distance, on average, of experimental data from the model predictions, and it has the same unit as the measurement. The results are summarized in Table 3 in terms of RMSE, where the subscripts “base” refers to the GPR developed for the old process, “migrate” refers the model for the new process via migration from the base GPR, and “new” corresponds to the model developed from the new process data only.

(Table 3 about here)

The results clearly indicate that the base model is not capable of predicting the new process satisfactorily. The migrated model presents more accurate predictions than the model built from the new process data only, especially when data from the new process are limited. Furthermore, when new process data are abundant, the migrated model appears to converge to the GPR using new data only, a desirable behavior for model migration algorithms.

Focusing on the case of adding 20 new design points, the difference between real and predicted production rates is shown in Figure 3. The vertical distance between the data and the diagonal line indicates prediction error. The figure clearly shows smaller prediction errors from the migrated model than those from the new model, especially when the production rate is greater than 200. In addition, Figure 4 shows the prediction errors (also known as residuals: $y_i - \hat{y}_i$). Since the residuals from the base model are very large, they are not shown in the figure for clearer illustration. The residual plot re-confirms the superior prediction accuracy of the migrated model.

(Figures 3 and 4 about here)

Next, the overall migration and optimization approach is applied to the CSTR example to maximize the production rate. The base GPR model, previously developed from 100 data points of the old process, is to be migrated to the new process for optimization purpose. Ten initial design points are allocated by the clustering method and the corresponding production rates are given in Table 4. Subsequently, EI is implemented to help allocate the next experiments for optimization. The entire optimization procedure is also included in Table 4. At the k -th run, a GPR is first migrated from the base model using the data from Run 1 to Run $k - 1$, followed by searching for the factors' value \mathbf{x} that gives the maximal EI (denoted by $\max(\text{EI})$ in Table 4). Subsequently, this \mathbf{x} is listed at the k -th run as the designed experiment to be conducted. The predicted production rate, \hat{y} , by the GPR is also given together with its 95% confidence bound. y_{real} is the real response obtained from simulation.

(Table 4 about here)

Run 8 presents the highest production rate of 1162.0 among the 10 initial experiments. By maximizing the measure of EI defined in Eq. (19), the next "experimental" run (No. 11) is allocated and given in the table. Note that the predicted production rate for the 11th run, 1196.8 ± 306.3 , is dramatically lower than the actual value of 1816.6. The poor prediction accuracy may be because of model extrapolation: the high production rate at Run 11 has never been observed in the initial 10 runs. Nevertheless, the measure of EI successfully identifies Run 11 as of potentially high production rate. Subsequently, the base GPR model is further migrated by using the 11 data points from the new process, and EI is again used to allocate Run 12. Compared with the previous run, the maximum of EI reduces slightly from 0.232 to 0.207, indicating a still high probability of further improving the process. As a consequence, Run 12 is experimented, giving rise to a production rate of 2252.7. Compared with the 11th run, the prediction for Run 12, 1888.4 ± 358.9 , is improved, though the prediction accuracy is still poor. However, judged by the criterion of serving the modeling purpose, i.e. the exploratory identification of process optimum that has not been observed before, the migrated GPR model should be viewed to perform satisfactorily by revealing potentially optimal process factors with very limited experimental data.

In principle, the model updating in conjunction with EI can be applied infinitely to guarantee to reach a global optimum [5]. In practice, after 12 runs, the maximal EI calculated based on the migrated GPR becomes a relatively small value (6.5×10^{-5}), suggesting little potential for subsequent improvement of production rate. Therefore, the iterative optimization procedure may be terminated after these 12 experiments. As a post-analysis, the obtained production rate of 2252.7 kmol/min can be verified by the simulation model to be the global optimum. Specifically, the global optimum was confirmed through constrained optimization (constraint being the factors range) based on the actual process model. To ensure global optimum, the optimization algorithm started from 1000 different initial values that are given by HSS to cover the entire design space.

Table 5 compares three strategies for optimizing the process. For a fair comparison, all strategies use 10 initial design points that are allocated by either clustering or HSS method. The proposed approach, "Clustering + Migration + EI", attains the best performance in terms of finding the global optimum with the fewest number of runs. By replacing clustering with HSS design, the initial DoE does not consider the information of process response that is available from the old and similar process (see the discussions in Section 2.3). As a result, the method of "HSS + Migration + EI" is trapped at a local optimum in this case study. Finally, the method of "HSS + EI" completely ignores the information of the old process; it essentially applies HSS for initial design of the new process. Then, this method iterates between GPR modeling using all data from the new process, and EI maximization to allocate and conduct a new experiment. Table 5 indicates that the method of "HSS + EI" only finds a local optimum after 37 runs in this example. Certainly, theoretical analysis shows that by using EI, the optimization algorithms (including "HSS + Migration + EI" and "HSS + EI") will ultimately find the global optimum, if an arbitrarily but often impractically large number of runs are allowed

[5]. In contrast, the proposed strategy effectively utilizes the information from the old process, and is capable of reaching the global optimum rapidly.

(Table 5 about here)

3.2. Catalytic epoxidation of *trans*-stilbene

The second example to validate the proposed model migration algorithm is related to a lab-scale catalytic reaction. Specifically, we are interested in maximizing the *trans*-stilbene conversion rate in the epoxidation of *trans*-stilbene over Co^{2+} -NaX catalyst. Previously, the epoxidation was investigated using molecular oxygen (air) as oxidant [9]. In this work, the use of tert-Butyl hydroperoxide (TBHP) as oxidant will be investigated. It appears that the two processes are very similar and the difference is only in the use of oxidant. However, the old process is a three-phase reaction (solid-phase catalysts, gas-phase oxygen and liquid-phase *trans*-stilbene) while the new process involves only two phases (TBHP is in liquid phase). As a result, the two processes are similar in the overall epoxidation reaction but significantly different in the underlying reaction mechanism.

Sodium form zeolite X (NaX) is purchased from Sigma-Aldrich. Unit cell composition of NaX is $\text{Na}_{88}\text{Al}_{88}\text{Si}_{104}\text{O}_{384}$ with unit cell size of 24.94Å. The BET surface area of the zeolite is $608 \text{ m}^2\text{g}^{-1}$. Cobalt-exchanged zeolite (Co^{2+} -NaX) is prepared by ion-exchange of the NaX with $0.1 \text{ mol}\cdot\text{L}^{-1}$ $\text{Co}(\text{NO}_3)_2$ aqueous solution with 1:80 ratio of NaX zeolite to $\text{Co}(\text{NO}_3)_2$. The reaction mixture is then heated at $80 \text{ }^\circ\text{C}$ for 4 h. The resulting powder is filtered and washed with deionized water to avoid un-exchanged cobalt ions. Finally, the washed Co^{2+} -NaX sample is dried at $100 \text{ }^\circ\text{C}$ for 4 h.

A batch-type reactor operated under atmospheric pressure is used to carry out the liquid phase catalytic *trans*-stilbene epoxidation reactions. In a typical reaction, a measured amount of *trans*-stilbene (>96%, Aldrich), 200 mg of Co^{2+} -NaX catalyst, and 10 mL of N,N-dimethylformamide (DMF, >99.8%, J.T.Baker) are added into a 50 mL round-bottomed flask. For the old process, O_2 or O_2 diluted with N_2 is bubbled into the liquid at a flow rate of $50 \text{ mL}\cdot\text{min}^{-1}$, whereas 10 mmol TBHP is added as oxidant in the new process. Time counting starts by immersing the round-bottomed flask into an oil bath under desired reaction temperature. The solid catalyst is filtered off after reaction. The liquid organic products were analyzed by an Agilent gas chromatograph (GC) 6890 equipped with a HP-5 capillary column (30 m long and 0.32 mm inner diameter, packed with silica-based supelcosil). Calibration of GC is performed with known amounts of benzaldehyde, benzoic acid, stilbene, and stilbene oxide in DMF. The conversion is obtained on the basis of moles of *trans*-stilbene as

$$\text{Conversion}(\%) = \frac{(\text{initial moles}) - (\text{final moles})}{(\text{initial moles})} \times 100\% \quad (26)$$

The old process was previously investigated [9] with five process factors being considered: reaction temperature, partial pressure of oxygen, initial concentration of *trans*-stilbene, stirring rate and reaction time. The process response that needs to be maximized is the conversion of *trans*-stilbene defined in Eq. (26). In the new process, partial pressure of oxygen is not considered since TBHP is used as oxidant. The range of the remaining four factors to be explored is listed in Table 6. The upper bound of temperature in the new process is lower since TBHP is volatile above $90 \text{ }^\circ\text{C}$.

(Table 6 about here)

The base GPR model was developed by using the 41 data points of the old process [9]. The clustering method is used to design six initial experiments for the new process. The number of initial experiments is partially determined by our experimental resources. The experimental results are summarized in Table 7. Among the initial six experiments, Run 4 gives the highest conversion of 78.8%.

(Table 7 about here)

The data from the initial experiments are used to migrate the base GPR model to the new process. Subsequently, the measure of EI is used to guide the search for higher conversion rate. The results are also listed in Table 7, where \hat{y} denotes the predicted conversion with corresponding 95% confidence bound and y_{real} is the actual conversion obtained from experiments. Similar to the simulated process in the previous sub-section, the first two predictions (Runs 7 and 8), given by GPR models that are migrated using only several data points from the new process, are not sufficiently accurate. The poor prediction accuracy also suggests that the 7th and 8th experiments are allocated to under-explored region of the factors’ space, which is a desired property of general iterative optimization algorithms. In addition, prior to conducting the 9th experiment, the $\max(\text{EI})$ maintains to be a significant value, indicating a high probability to obtain further improvement. Indeed, Run 9 presents a significant improvement in conversion (87.3%) over the previous eight experiments, and its response is well predicted by the GPR model ($\hat{y} = 86.6 \pm 10.3$). Continuing this iterative procedure, the maximal value of EI after Run 9 drops dramatically to 9.03×10^{-4} , suggesting a low probability of obtaining any further progress. In addition, the criterion of $\max(\text{EI})$ allocates the next experiment that has very similar process factors to Run 9 (the only difference between Runs 9 and 10 is the 1 °C difference in reaction temperature), which is also an indication that the algorithm has reached the process optimum. Nevertheless, as a confirmation experiment, Run 10 is conducted anyway and the resultant conversion is 87.6%, only slightly higher than that of Run 9. After Run 10, the maximal value of EI decreases to be close to zero (5.82×10^{-16}), and thus the iterative optimization algorithm is terminated.

Recall that for the old process, a total of 41 experiments were conducted to maximize the conversion rate. With the model migration and iterative optimization approach, the optimal condition for a new yet similar process has been obtained with only 10 experiments. This comparison, although not theoretically rigorous, indicates that the proposed approach significantly reduces the number of experiments required to reach a process optimum for a similar process through effective utilization of the existing information of the old process.

4. Concluding remarks

This paper presents an integrated model migration and model-based optimization method for rapidly identifying the process optimum with limited experiments. The proposed method is an extension of the previous “two-step” approach whereby the model is first migrated from a old process to the new process with sufficient prediction accuracy, prior to being used for optimization purpose. Each component of the integrated framework has been carefully chosen, including (i) clustering as initial DoE make the most use of the information from the old process, (ii) a GPR as empirical process model and the corresponding Bayesian migration algorithm to fully account for model uncertainty, and (iii) a statistical measure of expected improvement (EI) to guide the optimizing search in the presence of model uncertainty. The proposed methodology has been applied to the optimization of a simulated chemical process, and a real catalytic reaction for the epoxidation of *trans*-stilbene.

The prediction from the empirical model appears to be inaccurate in the case studies. This phenomenon is mainly due to the lack of sufficient experimental data and thus the model prediction amounts to extrapolation. In this regard, the proposed methodology may be viewed as a powerful tool to qualitatively indicate the potential direction for process improvement. If experimental resources allow, the obtained process optimum may be confirmed/refined by conducting more iterations of the proposed method. Theoretical analysis showed that global optimum is guaranteed in the limit of infinite number of iterations [5].

In this work, the presented method is limited to allocate a single experiment at each iteration. While this is cost-effective in many applications, it may be desired to design multiple experiments at each iteration when, e.g., a parallel automatic reaction system is available for high-throughput experimentation [36]. A possible solution is to search for all local maxima of the EI as candidates for experiments. If too many local maxima exist, then a usual clustering method can be adopted to extract the most informative maxima as the design points. Significant work remains to be conducted in this area in the future.

In addition, the experimental work presented in this study is based on a lab-scale reactor where the process factors can be well controlled. When the proposed method is applied to more practical situations where process factors cannot be precisely manipulated, robust design and optimization approach may be needed to minimize the process variability. The combination of robust design and optimization approach with the proposed modeling and optimization method will be further investigated.

Acknowledgment

Financial support from Singapore AcRF Tier 1 grant (RG 19/09) is acknowledged. Xinni Lin participated in the catalytic epoxidation experiments as a partial requirement of her final year project.

References

- [1] K. Klatt, W. Marquardt, Perspectives for process systems engineering: Personal views from academia and industry, *Computers and Chemical Engineering* 33 (2009) 536–550.
- [2] R. H. Myers, D. C. Montgomery, *Response Surface Methodology: Process and Product in Optimization Using Designed Experiments*, John Wiley and Sons, Inc., 1995.
- [3] M. A. Karimi, H. Karami, M. Mahdipour, ANN modeling of water consumption in the lead-acid batteries, *Journal of Power Sources* 172 (2007) 946–956.
- [4] A. Niazi, S. Sharifi, E. Amjadi, Least-squares support vector machines for simultaneous voltammetric determination of lead and tin: A comparison between LS-SVM and PLS in voltammetric data, *Journal of Electroanalytical Chemistry* 623 (2008) 86–92.
- [5] D. R. Jones, A taxonomy of global optimization methods based on response surfaces, *Journal of Global Optimization* 21 (2001) 345–383.
- [6] A. O’Hagan, Curve fitting and optimal design for prediction, *Journal of the Royal Statistical Society B* 40 (1978) 1–42.
- [7] C. E. Rasmussen, C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.
- [8] J. Sacks, W. Welch, T. Mitchell, H. Wynn, Design and analysis of computer experiments, *Statistical Science* 4 (1989) 409–423.
- [9] Q. Tang, Y. Lau, S. Hu, W. Yan, Y. Yang, T. Chen, Response surface methodology using Gaussian processes: Towards optimizing the trans-stilbene epoxidation over Co^{2+} -NaX catalysts, *Chemical Engineering Journal* 156 (2010) 423–431.
- [10] J. Yuan, K. S. Wang, T. Yu, M. L. Fang, Reliable multi-objective optimization of high-speed WEDM process based on Gaussian process regression, *International Journal of Machine Tools and Manufacture* 48 (2008) 47–60.
- [11] J. D. Lu, K. Yao, F. R. Gao, Process similarity and developing new process models through migration, *AIChE Journal* 55 (2009) 2318–2328.
- [12] J. Lu, F. Gao, Process modeling based on process similarity, *Industrial and Engineering Chemistry Research* 47 (2008) 1967–1974.
- [13] J. D. Lu, Y. Yao, F. R. Gao, Model migration for development of a new process model, *Industrial and Engineering Chemistry Research* 48 (2009) 9603–9610.
- [14] J. D. Lu, F. R. Gao, Model migration with inclusive similarity for development of a new process model, *Industrial and Engineering Chemistry Research* 47 (2008) 9508–9516.

- [15] V. Chen, K. Tsui, R. Barton, M. Meckesheimer, A review on design, modeling and applications of computer experiments, *IIE Transactions* 38 (2006) 273–291.
- [16] T. Chen, J. Morris, E. Martin, Gaussian process regression for multivariate spectroscopic calibration, *Chemometrics and Intelligent Laboratory Systems* 87 (2007) 59–67.
- [17] Z. G. Qian, C. C. Seepersad, V. R. Joseph, J. K. Allen, C. F. J. Wu, Building surrogate models based on detailed and approximate simulations, *Journal of Mechanical Design* 128 (2006) 668–677.
- [18] J. Lin, S. Jang, Nonlinear dynamic artificial neural network modeling using an information theory based experimental design approach, *Industrial and Engineering Chemistry Research* 37 (1998) 3640–3651.
- [19] J. Chen, D. Wong, S. Jang, S. Yang, Product and process development using artificial neural-network model and information analysis, *AIChE Journal* 44 (1998) 876–887.
- [20] M. Coleman, D. Block, Nonlinear experimental design using Bayesian regularized neural networks, *AIChE Journal* 53 (2007) 1496–1509.
- [21] K. T. Fang, D. K. J. Lin, P. Winker, Y. Zhang, Uniform design: Theory and application, *Technometrics* 42 (2000) 237–248.
- [22] M. D. McKay, R. J. Beckman, W. J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* 42 (2000) 55–61.
- [23] K. T. Fang, Y. Wang, P. M. Bentler, Some application of number-theoretic methods in statistics, *Statistical Science* 9 (1994) 416–428.
- [24] J. R. Kalagnanam, U. M. Diwekar, An efficient sampling technique for off-line quality control, *Technometrics* 39 (1997) 308–319.
- [25] T. Chen, B. Wang, Bayesian variable selection for Gaussian process regression: Application to chemometric calibration of spectrometers, *Neurocomputing* 73 (2010) 2718–2726.
- [26] F. di Sciascio, A. N. Amicarelli, Biomass estimation in batch biotechnological processes by Bayesian Gaussian process regression, *Computers and Chemical Engineering* 32 (2008) 3264–3273.
- [27] J. Q. Shi, B. Wang, R. Murray-Smith, D. M. Titterton, Gaussian process functional regression modeling for batch data, *Biometrics* 63 (2007) 714–723.
- [28] P. Zhou, F. F. Tian, X. Chen, Z. C. Shang, Modeling and prediction of binding affinities between the human amphiphysin sh3 domain and its peptide ligands using genetic algorithm-gaussian processes, *Biopolymers* 90 (2008) 792–802.
- [29] A. F. Hernandez, M. A. Grover, Stochastic dynamic predictions using kriging for nanoparticle synthesis, in: R. M. de Brito Alves, C. A. O. do Nascimento, E. C. Biscaia Jr. (Eds.), 10th International Symposium on Process Systems Engineering: Part A, Vol. 27 of Computer Aided Chemical Engineering, Elsevier, 2009, pp. 357 – 362.
- [30] M. Alvarez, N. Lawrence, Sparse convolved Gaussian processes for multi-output regression, in: D. Koller, D. Schuurmans, Y. Bengio, L. Bottou (Eds.), *Advances in Neural Information Processing Systems (NIPS)*, Vol. 21, 2009, pp. 57–64.
- [31] S. L. Chiu, Fuzzy model identification based on cluster estimation, *Journal of Intelligent and Fuzzy Systems* 2 (1994) 267–278.

- [32] R. Xu, D. Wunsch, Survey of clustering algorithms, *IEEE Transactions on Neural Networks* 16 (2005) 645–678.
- [33] M. Kennedy, A. O’Hagan, Bayesian calibration of computer models, *Journal of the Royal Statistical Society Series B* 63 (2001) 425–450.
- [34] M. Frean, P. Boyle, Using Gaussian processes to optimize expensive functions, in: W. Wobcke, M. Zhang (Eds.), *AI 2008: Advances in Artificial Intelligence*, Vol. 5360 of *Lecture Notes in Computer Science*, 2008, pp. 258–267.
- [35] U. M. Diwekar, E. S. Rubin, Parameter design methodology for chemical processes using a simulator, *Industrial and Engineering Chemistry Research* 33 (1994) 292–298.
- [36] L. Baumes, D. Farrusseng, M. Lengliz, C. Mirodatos, Using artificial neural networks to boost high-throughput discovery in heterogeneous catalysis, *QSAR and Combinatorial Science* 23 (2004) 767–778.

Table 1: Reaction kinetic parameters (k_A and k_B) in the Arrhenius equation $k = A \exp(-E_a/RT)$ for the old and new processes.

Parameter	Old process		New process	
	A	E_a	A	E_a
k_A	8.4×10^5	3.64×10^4	1.0×10^6	3.04×10^4
k_B	7.6×10^4	3.46×10^4	6.0×10^4	3.66×10^4

Table 2: Range of the process factors in the CSTR simulation.

Factor	Range in the old process	Range in the new process
C_{Ai} (mol/m ³)	1000 – 5000	800 – 4000
C_{Bi} (mol/m ³)	100 – 500	100 – 500
T (K)	290 – 330	290 – 380
V (m ³)	10 – 90	10 – 90

Table 3: Comparison of prediction performance in terms of RMSE with different number of data points from the new process.

Method	No. of data points					
	5	10	20	30	50	80
RMSE _{migrate}	133.4	73.9	34.1	28.0	5.7	4.2
RMSE _{new}	159.5	130.2	55.2	28.7	5.9	4.2
RMSE _{base}	288.5	291.3	286.7	271.4	280.0	274.0

Table 4: Results of iterative maximization of the production rate.

Run No.	C_{Ai}	C_{Bi}	T	V	y_{real}	\hat{y}	max(EI)
1	1664.0	271.9	341.1	27.9	93.8	–	–
2	2656.0	268.8	355.6	47.8	392.2	–	–
3	1920.0	296.9	314.4	79.1	112.4	–	–
4	3968.0	300.0	350.0	74.0	687.5	–	–
5	3072.0	212.5	318.9	25.4	60.3	–	–
6	2240.0	131.3	347.8	85.5	402.4	–	–
7	1120.0	359.4	376.3	78.5	646.4	–	–
8	3040.0	312.5	368.9	86.2	1162.0	–	–
9	864.0	390.6	296.3	27.3	11.6	–	–
10	3136.0	162.5	378.9	57.4	986.8	–	–
11	3093.4	500.0	380.0	88.9	1816.6	1196.8±306.3	0.232
12	4000.0	500.0	380.0	90.0	2252.7	1888.4±358.9	0.207
–	3744.0	484.4	291.1	84.9	–	529.5±1008.8	6.5×10^{-5}

Table 5: Comparison of different methods for maximizing the production rate.

Method	No. of runs	Identified optimum	Global optimum?
Clustering + Migration + EI	12	2252.7	Yes
HSS + Migration + EI	13	2028.7	No
HSS + EI	37	1156.8	No

Table 6: Range of process factors in the epoxidation process.

Process factor	Range of the old process	Range of the new process
Temperature, T (°C)	60 – 120	40 – 90
Initial stilbene concentration, c (mmol/15 mL)	1 – 5	1 – 5
Stirring rate, $stir$ (rpm)	200, 300, 400, 500, 700, 1000, 1250	200, 300, 400, 500, 700, 1000, 1250
Reaction time, t (min)	30 – 240	30 – 240

Table 7: Results of iterative maximization of the conversion during the epoxidation of *trans*-stilbene.

Run No.	T	c	$stir$	t	$y_{real}(\%)$	$\hat{y}(\%)$	max(EI)
1	86	3.93	200	185	60.3	–	–
2	53	3.50	500	138	24.2	–	–
3	84	2.26	1250	129	63.9	–	–
4	86	1.45	200	214	78.8	–	–
5	72	2.25	700	40	17.9	–	–
6	90	1.00	1250	120	66.0	–	–
7	90	1.00	1250	240	77.8	103.5±8.5	1.067
8	90	3.90	1250	240	57.9	87.6±3.2	0.429
9	76	1.00	1250	240	87.3	86.6±10.3	0.510
10	77	1.00	1250	240	87.6	87.0±1.2	9.03×10^{-4}
–	90	1.94	200	222	–	77.8±8.6	5.82×10^{-16}

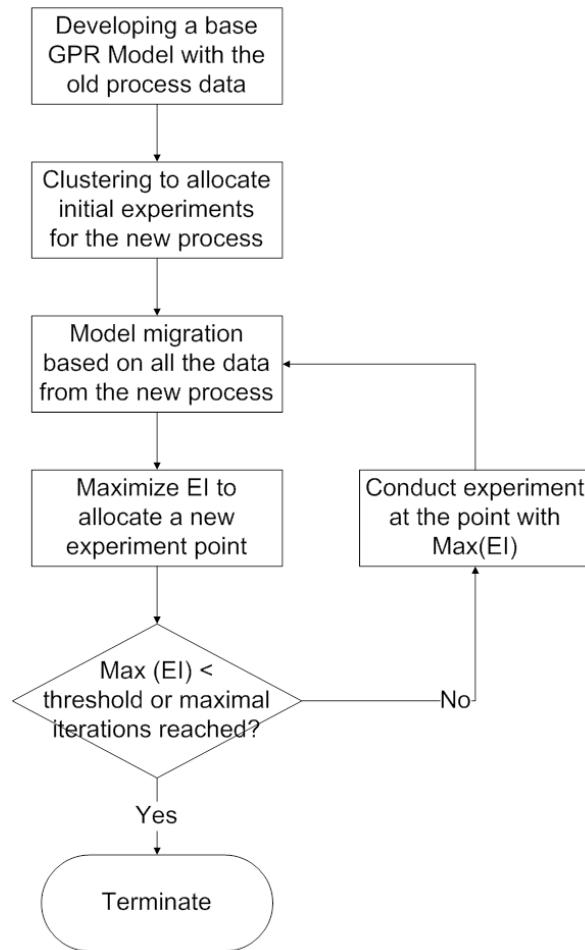


Figure 1: Model migration flowchart and process optimization method.

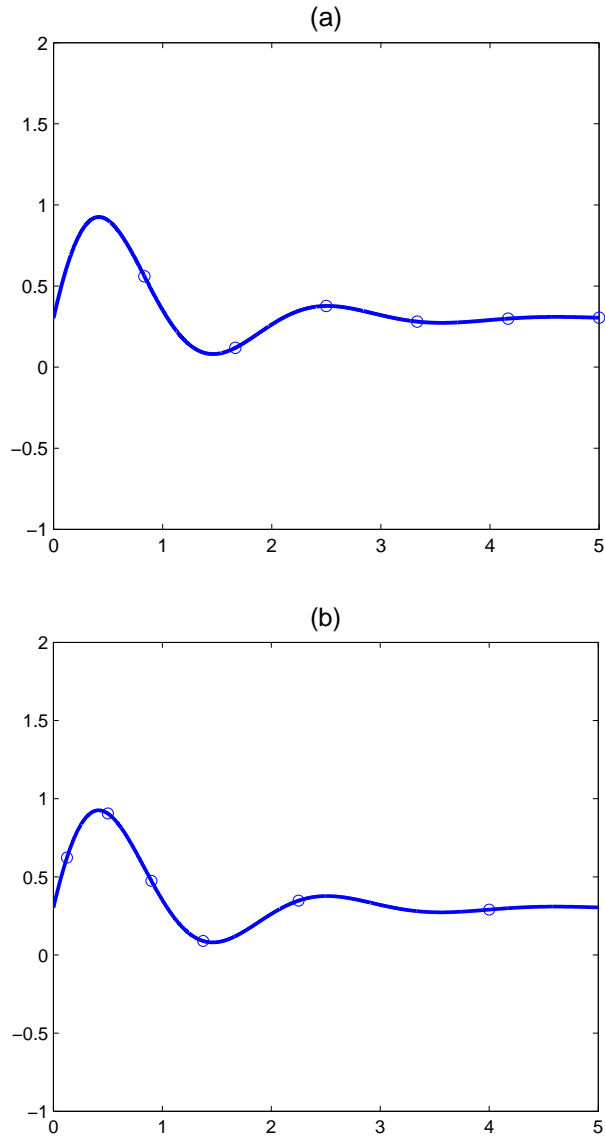


Figure 2: DoE method (a) HSS to allocate six points uniformly distributed; (b) DoE by considering both x and y but not uniformly distributed.

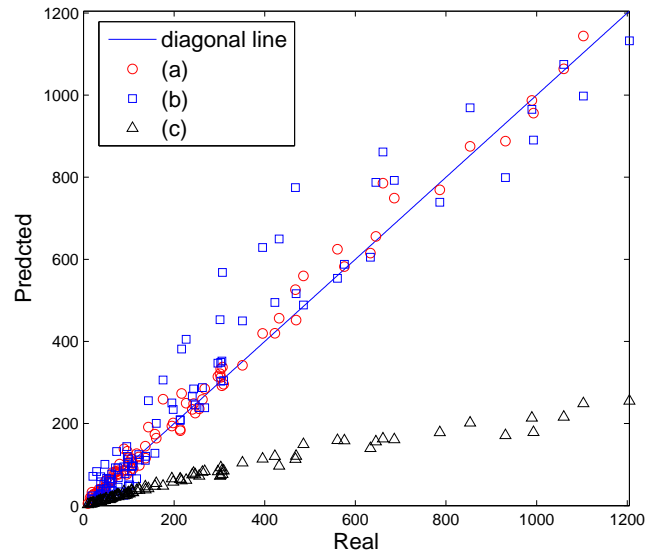


Figure 3: Prediction performance of (a) the migrated model (RMSE=34.1); (b) model using the new process data only (RMSE=55.2); (c) the base model (RMSE=286.7).

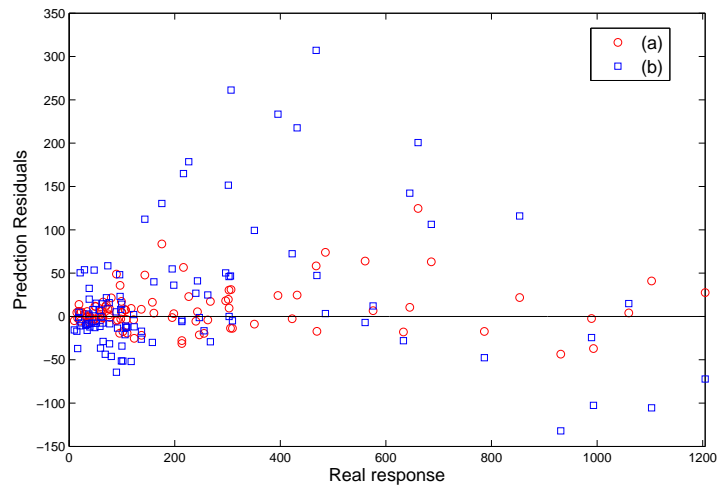


Figure 4: Prediction residuals of (a) the migrated model (RMSE=34.1); (b) model using the new process data only (RMSE=55.2).