Preprints of the
9th International Symposium on Advanced Control of Chemical Processes
The International Federation of Automatic Control
June 7-10, 2015, Whistler, British Columbia, Canada



# Model Migration through Bayesian Adjustments

**Linkai, Luo** * **Furong, Gao** **

* *Department of Chemical & Biomoleculor Engineering,The Hong Kong University of Science & Technology, Hong Kong (e-mail: lkluo@ust.hk)*
** *Department of Chemical & Biomoleculor Engineering,The Hong Kong University of Science & Technology, Hong Kong (e-mail: kefg@ust.hk)*

**Abstract:**
Model migration has been proved to be an effective modeling tool to adopt an existing base model from an old process to a similar, yet non-identical process. However, if the process differences are more complex and differ from sample to sample, then the existing model migration strategies can be non-flexible and inadequate. Based on the concepts laid out in an earlier article (Lu and Gao (2008b)), this paper presents an enhanced Bayesian model migration strategy for statistical models. This is achieved by applying Bayesian adjustments to a base model developed using the Gaussian process (GP). The benefits of the proposed method are demonstrated on a continuously stirred tank reactor.

*Keywords:* Process models, Gaussian processes, parameter estimation, optimal experimental design, efficient algorithms

## 1. INTRODUCTION

Models which describe the physical processes under investigation are essential for quality prediction, model-based process design, control and optimization. One approach to develop models is based on physical laws that dictate the model structures. However, developing suck kind of models commonly requires considerable experience and computation. An alternative approach to make better predictions within the processes is to use the statistics techniques to postulate models based on the experimental data observed from the real processes. Typical examples include polynomial regression, artificial neural networks (ANNs), Gaussian process (GP) models, and so on. Since statistical models (or empirical models in some literatures) are advantageous to first-principle models in terms of simplicity and being able to provide good predictions when the data points are close to the observed training data, they are usually adopted to processes where theatrical models are not readily available.

This paper is focused on such scenarios to develop statistical models for certain processes. While there is often mature technique for a variety of statistical models, their development can be quite time-consuming and expensive due to a large amount of required data. This is particularly relevant in areas where processes are operated over a range of conditions to produce various products. For instance, if the production of a certain grade is transferred to a second process, a full redevelopment procedure is necessary for the changed process. On the other hand, although processes fashioning products may differ in size, configuration, or use different technology, there is evidence that some of them share similar mechanisms. In this context, by borrowing strength from an existing model for use in the development of a similar, yet non-identical process, it is possible to avoid those time-consuming redevelopment procedures. Some has been published on this topic. One of those examples is calibration transfer between two or several near infrared (NIR) instruments (Fearn (2001) and Feudale et al. (2002)). The model is developed on a first spectrometer called "*master*" instrument and is used on a second or several other spectrometers, i.e., "*slave*" devices via appropriate adjustment. Most of the calibration models are linear in terms of partial least squares (PLS) and principal component regression (PCR). As such, those methods may not be suitable for nonlinear programs. Another example is model migration proposed by Lu and Gao (2008b). In their work, *base* model is built based on an *old* process and then is migrated to the *new* model associated with a similar, yet non-identical *new* process. According to Lu et al. (2009), similar processes can be identified by either descriptive attributes that belong to a pair or several processes, or quantitative models. They often arise when qualitative changes are made across processes. For example, injection molding processes that make similar polymer products with various shapes and materials can be considered similar processes, see Lu et al. (2009). The benefits of model migration can be summarized as follows. With comparable prediction accuracy, fewer data are needed for developing new model through migration strategies than those for creating the base model. Thus, model migration is a collection of efficient modeling methods which incorporate a few data points from new processes with base models.

The purpose of this paper is to provide new viewpoints on model migration originated from Lu and Gao (2008b). Other related work includes Lu et al. (2008) and Lu and Gao (2008a). As noted from the work of Lu and Gao (2008b) , new model prediction can be formulated by correcting predicted values from base model through input-output slope and bias correction (SBC) method:

$$y_{new} = S_O f(S_I \boldsymbol{x}_{new} + B_I) + B_O, \quad (1)$$

where $f(\cdot)$ denotes base model of any form, $S$ and $B$ are slope and bias parameters,respectively, and the subscripts $I$ and $O$ are shorthands for input and output, respectively. The newly introduced slope/bias parameters can be estimated using least squares method, together with a few training data experimented from new process. SBC works well when the process differences are simple and systematic in all data observed from new process; however, if the differences are more complex and differ

from sample to sample, such migration strategy seems likely non-flexible. In addition, the parameter estimates will pursuit stable as training data grow (it will be demonstrated in Section 5). As a result, the original SBC is not ideally suitable for model enhancement when there are abundant training data. Yan et al. (2011) extended previous SBC to functional SBC, as they noticed that the previous parametric SBC updating rule may not possess sufficient flexibility to model the process of interest. Also they argued that the original SBC would ultimately result in a linear model if the base model was developed with linear regression methods. As a result, the parametric SBC can not offer the flexibility to account for new processes that are nonlinear. In the work of Yan et al. (2011), they proposed to applying Gaussian process (GP) to migration, i.e., replacing slope/bias parameters with GP functions. Their method enhanced migration by borrowing strength of the GP technique, but limited to the correction for the processes' output. As pointed out in Lu et al. (2008) and Vastola et al. (2013), correction in inputs of similar, yet non-identical processes is also necessary to account for more complex process differences, because the corresponding points for similar, yet non-identical processes may not be identical. To this end, careful adjustments need to make to address the defect that may arise in the literatures, and form the primary focus of our work.

This paper looks at model migration through Bayesian adjustments. The use of Bayesian adjustments for integrating multiple sources is not a new concept. For example, Qian and Wu (2008) introduced Bayesian hierarchical Gaussian process models to integrate *low-accuracy experiment* and *high-accuracy experiments* efficiently. Their method was based on a fully Bayesian approach that assumed priors for all model parameters. To estimate the parameters and make predictions, it required a Markov Chain Monte Carlo (MCMC) simulation that may be computationally prohibitive in some problems. To ease the computational burden, we propose a Bayesian adjustments approach to model migration which can provide a closed-form expression of the predictive distributions and account for uncertainties in the model parameters. Additional, we suggest to employ GP models for use in the development of an old process and model migration, since GP models have the advantage of being able to accurately model non-linear and complex functions that frequently occur in science and engineering (see Rasmussen (2006)).

After a brief overview of GP models in Section 2, we discuss Bayesian adjustments for model migration in Section 3, followed by considerations on how to plan experiments for migration purpose in Section 4. A continuously stirred tank reactor (CSTR) is used to demonstrate and compare the Bayesian adjustments against other modeling strategies in Section 5. Section 6 discusses the remaining issues related to the current work and summaries the conclusions.

## 2. DEVELOPING BASE MODELS USING GP

GP models have been extensively applied to process modeling in science and engineering because of their capability of accurately representing complex and nonlinear relationships within the processes. In general, GP is a collection of stochastic random variables over functions, any finite number of which have a joint Gaussian distribution ( Rasmussen (2006)). Let $\boldsymbol{x} = (x_1, \ldots, x_d)$ denote a $d$-dimensional input factors, and the matrix $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$, of size $n \times d$, all $n$ training input vectors. For the variable outputs $\boldsymbol{y} = (y_1, \ldots, y_n)^T$, GP models the regression function having a Gaussian prior distribution with zero mean and certain covariance, i.e.,

$$\boldsymbol{y} \sim \mathsf{N}(0, \boldsymbol{C} + \sigma^2 \boldsymbol{I}), \tag{2}$$

where $\boldsymbol{C}$ is the $n \times n$ covariance matrix with elements defined by a covariance function, $C_{ij} = C(\boldsymbol{x}_i, \boldsymbol{x}_j)$, and $\sigma^2$ is the noise variance and $\boldsymbol{I}$ is the identity matrix. The choice of the covariance function is nontrivial. Here and throughout this work, we use the stationary square exponential covariance function, which produces smooth functions, given by

$$\begin{aligned} C(\boldsymbol{x}_i, \boldsymbol{x}_j) &= \sigma_\ell^2 R(\boldsymbol{x}_i, \boldsymbol{x}_j) \\ &= \sigma_\ell^2 \exp\big[-\frac{1}{2}\sum_{k=1}^{d} \ell_k^{-2}(x_{ik} - x_{jk})^2\big], \end{aligned} \tag{3}$$

where $\sigma_\ell$ and $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_d)$ are the hyper-parameters of the GP model.

The hyper-parameters $\boldsymbol{\theta} = (\sigma_\ell, \boldsymbol{\ell}, \sigma)$ can be estimated by maximization of the marginal likelihood $\log p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\theta})$:

$$-\frac{1}{2}\boldsymbol{y}^T(\boldsymbol{C} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{y} - \frac{1}{2}\log|\boldsymbol{C} + \sigma^2\boldsymbol{I}| - \frac{1}{2}\log(2\pi). \tag{4}$$

This optimization problem can be solved using the derivative of the log-likelihood with respect to each hyper-parameter. Given the training data $\boldsymbol{X}$ and $\boldsymbol{y}$, together with the resulting hyper-parameters $\boldsymbol{\theta}$, the conditional predictive distribution of $y_* \equiv y(\boldsymbol{x}_*)$ for a new data point $\boldsymbol{x}_*$ is again Gaussian with mean $\bar{y}_*$ and variance $V_*$:

$$\bar{y}_* = \boldsymbol{c}^T(\boldsymbol{C} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{y} \tag{5}$$

$$V_* = \boldsymbol{c}_* - \boldsymbol{c}^T(\boldsymbol{C} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{c}, \tag{6}$$

where $\boldsymbol{c} = [C(\boldsymbol{x}_*, \boldsymbol{x}_1), \ldots, C(\boldsymbol{x}_*, \boldsymbol{x}_n)]^T$, and $\boldsymbol{c}_* = C(\boldsymbol{x}_*, \boldsymbol{x}_*)$.

## 3. MODEL MIGRATION

In the older version of slope/bias correction discussed by Lu and Gao (2008b), the parameters seem likely constant for all the input variables. Such correction appears to be insufficient if the variable magnitudes vary significantly. Although the input data can be normalized to the same range, the impact of slope/bias parameters shall be independent among each other. We consider our model migration method with the existence of a relationship between similar processes as follows:

(1) Each input variable of old process corresponds to that of new process through the following equation:

$$\tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i\boldsymbol{\rho}_X + \boldsymbol{\lambda}_X.$$

where $\tilde{\boldsymbol{x}}$ denotes the input variables of old process, and $\boldsymbol{x}$ denotes input variables of new process. The matrix $\boldsymbol{\rho}_X = \mathsf{diag}(\rho_{X,1}, \ldots, \rho_{X,d})$ contains slope parameters in input variables, and the vector $\boldsymbol{\lambda}_X = (\lambda_{X,1}, \ldots, \lambda_{X,d})$ contains bias parameter, accordingly. In order to eliminate an exhaustive treatment on these parameters, Vastola et al. (2013) suggests to use $\rho_{X,k} = 1, k = 1, \ldots, d$, if the process difference are simple. Note that the translation is invertible.

(2) The predictive output, $\tilde{y}_i = f(\tilde{\boldsymbol{x}}_i)$, from base model will be augmented by introducing slope/bias parameters $\rho_Y$ and $\lambda_Y$ to best reflect the expected result of the new model:

$$y_i = \rho_Y \cdot \tilde{y}_i + \lambda_Y + \varepsilon_i.$$

where $\varepsilon \sim \mathsf{N}(0, \sigma_\varepsilon^2)$ is the measurement error.

We put the aforementioned procedures together to yield a reformulated migration strategy:

$$y_i = \rho_Y \cdot f(\boldsymbol{x}_i\boldsymbol{\rho}_X + \boldsymbol{\lambda}_X) + \lambda_Y + \varepsilon_i. \tag{7}$$

In (7), the parameters of interest are $\boldsymbol{\rho}_X$, $\boldsymbol{\lambda}_X$, $\rho_Y$ and $\lambda_Y$, which need to be estimated using observed data from a new process. Given $m$ data points, $\boldsymbol{X}_m = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)^T$ and $\boldsymbol{y} = (y_1, \ldots, y_m)^T$, where $m < n$, we can find $\boldsymbol{\rho}_X$, $\boldsymbol{\lambda}_X$, $\rho_Y$ and $\lambda_Y$ by minimizing the sum of squared prediction errors:

$$\arg\min \sum_{k=1}^{m} \varepsilon_i^2, \tag{8}$$

subject to

$$\varepsilon_i = y_i - [\rho_Y \cdot f(\boldsymbol{x}_i \boldsymbol{\rho}_X + \boldsymbol{\lambda}_X) + \lambda_Y].$$

It has been argued that model migration (7) is non-flexible and insufficient to account for nonlinear relationship that may exist in a new process. To address this issue, slope/bias functions are introduced to replace the constants in (7), i.e.,

$$y(\boldsymbol{x}_i) = \rho_Y \cdot f(\boldsymbol{x}_i \boldsymbol{\rho}_X + \boldsymbol{\lambda}_X) + \lambda_Y(\boldsymbol{x}_i) + \varepsilon_i. \tag{9}$$

The bias correction function $\lambda_Y(\boldsymbol{x})$ is assumed to be GP taking the form

$$\lambda_Y(\boldsymbol{x}) = \bar{\lambda}_Y(\boldsymbol{x}) + Z(\boldsymbol{x}), \tag{10}$$

where $\bar{\lambda}_Y(\boldsymbol{x})$ is the mean function, and $Z(\boldsymbol{x})$ is GP having mean zero and covariance function $C$ defined as in (3). A commonly choice of $\bar{\lambda}_Y(\boldsymbol{x})$ is a first-order linear regression

$$\bar{\lambda}_Y(\boldsymbol{x}) = \lambda_0 + \sum_{k=1}^{d} \lambda_k x_k, \tag{11}$$

where $\boldsymbol{\lambda} = (\lambda_0, \ldots, \lambda_d)$ is the unknown parameters vector. The inference of model (9) is non-trivial after adding a stochastic term $\lambda_Y(\boldsymbol{x})$. One can, of course, use a numerical method, e.g., MCMC, to solve the nonlinear problem; however, such an attempt is computationally unfriendly. An alternative way of handling computation burden and of providing a close-form analytical expression of the predictive distribution is described as follows. In the first stage, we obtain estimate of the slope/bias parameter in input variables, $\hat{\boldsymbol{\rho}}_X$, $\hat{\boldsymbol{\lambda}}_X$, by optimization (8). In the second stage, we use those estimates to replace the unknown quantities, i.e., the migration (9) then becomes

$$y(\boldsymbol{x}_i) = \rho_Y \cdot f(\boldsymbol{x}_i \hat{\boldsymbol{\rho}}_X + \hat{\boldsymbol{\lambda}}_X) + \lambda_0 + \sum_{k=1}^{d} \lambda_k x_k + Z(\boldsymbol{x}_i) + \varepsilon_i. \tag{12}$$

Model (12) is easier to infer compared to model (9). A Bayesian strategy can be adopted via appropriate prior. We begin with writing model (12) in a matrix form

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{\beta} + \boldsymbol{Z} + \boldsymbol{\varepsilon}, \tag{13}$$

where

$$\boldsymbol{H} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1d} & f(\boldsymbol{x}_1 \hat{\boldsymbol{\rho}}_X + \hat{\boldsymbol{\lambda}}_X) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{m1} & \cdots & x_{md} & f(\boldsymbol{x}_m \hat{\boldsymbol{\rho}}_X + \hat{\boldsymbol{\lambda}}_X) \end{bmatrix}$$

is the regression matrix, $\boldsymbol{\beta} = (\lambda_0, \lambda_1, \ldots, \lambda_d, \rho_Y)^T$ is the collection of regression parameters, $\boldsymbol{Z} = (Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_m))^T$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_m)^T$. Note that $\boldsymbol{Z} + \boldsymbol{\varepsilon} \sim \mathsf{N}(0, \sigma_\ell^2 \boldsymbol{R}_\varepsilon)$, where $\boldsymbol{R}_\varepsilon = \boldsymbol{R} + \tau \boldsymbol{I}$, $\tau = \frac{\sigma_\varepsilon^2}{\sigma_\ell^2}$ and $\boldsymbol{R}$ is the $m \times m$ correlation matrix having $ij$-th element $R(\boldsymbol{x_i}, \boldsymbol{x}_j)$. According to (13), the log likelihood (ignoring constant terms) of the new model observations given base model is

$$-\frac{m}{2}\log(\sigma_\ell^2) - \frac{1}{2}\log|\boldsymbol{R}_\varepsilon| - \frac{(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{\beta})^T \boldsymbol{R}_\varepsilon^{-1} (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{\beta})}{2\sigma_\ell^2}. \tag{14}$$

We are interested in the predictive distribution of the new model at a testing point $\boldsymbol{x}_*$. Instead of applying a point estimation to $\boldsymbol{\beta}$, we use Bayes theorem to integrate $\boldsymbol{\beta}$ out to obtain the predictive distribution

$$p(y_*|\boldsymbol{y}) = \int p(y_*|\boldsymbol{y}, \boldsymbol{\beta}) p(\boldsymbol{\beta}|\boldsymbol{y}) \mathrm{d}\boldsymbol{\beta}, \tag{15}$$

where the first term in the integral is the conditional distribution of the response data given parameters $\boldsymbol{\beta}$ and training data $\boldsymbol{y}$, which is also Gaussian distributed according to Bastos and O'Hagan (2009). The second term $p(\boldsymbol{\beta}|\boldsymbol{y})$ is the posterior distribution of the regression parameters and is therefore calculated via Bayes theorem: $p(\boldsymbol{\beta}|\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{\beta}) p(\boldsymbol{\beta}) / p(\boldsymbol{y})$, where $p(\boldsymbol{y}|\boldsymbol{\beta})$ is the likelihood function given data, $p(\boldsymbol{\beta})$ is the prior distribution and $p(\boldsymbol{y})$ is a normalized constant. Using a weak prior for $\boldsymbol{\beta}$, $p(\boldsymbol{\beta}) \propto 1$, and combing with Bayes theorem, it can be shown that

$$\boldsymbol{\beta}|\boldsymbol{y} \sim \mathsf{N}\big(\hat{\boldsymbol{\beta}}, \sigma_\ell^2 (\boldsymbol{H}^T \boldsymbol{R}_\varepsilon^{-1} \boldsymbol{H})^{-1}\big), \tag{16}$$

where $\hat{\boldsymbol{\beta}} = (\boldsymbol{H}^T \boldsymbol{R}_\varepsilon^{-1} \boldsymbol{H})^{-1} \boldsymbol{H}^T \boldsymbol{R}_\varepsilon^{-1} \boldsymbol{y}$. Integrating out $\boldsymbol{\beta}$ in (15), it can be shown that

$$y_*|\boldsymbol{y} \sim \mathsf{N}(\bar{y}_*, V_*) \tag{17}$$

where

$$\bar{y}_* = \boldsymbol{h}(\boldsymbol{x}_*)^T \hat{\boldsymbol{\beta}} + \boldsymbol{r}_\varepsilon(\boldsymbol{x}_*)^T \boldsymbol{R}_\varepsilon^{-1} (\boldsymbol{y} - \boldsymbol{H}\hat{\boldsymbol{\beta}})$$

$$V_* = \sigma_\ell^2 \Big[ 1 - \boldsymbol{r}_\varepsilon(\boldsymbol{x}_*)^T \boldsymbol{R}_\varepsilon^{-1} \boldsymbol{r}_\varepsilon(\boldsymbol{x}_*) + \big(\boldsymbol{h}(\boldsymbol{x}_*) - \boldsymbol{r}_\varepsilon(\boldsymbol{x}_*)^T \boldsymbol{R}_\varepsilon^{-1} \boldsymbol{H}\big) \cdot \big(\boldsymbol{H}^T \boldsymbol{R}_\varepsilon^{-1} \boldsymbol{H}\big)^{-1} \big(\boldsymbol{h}(\boldsymbol{x}_*) - \boldsymbol{r}_\varepsilon(\boldsymbol{x}_*)^T \boldsymbol{R}_\varepsilon^{-1} \boldsymbol{H}\big)^T \Big],$$

with $\boldsymbol{h}(\boldsymbol{x}) = [1, x_1, \ldots, x_d, f(\boldsymbol{x}\hat{\boldsymbol{\rho}}_X + \hat{\boldsymbol{\lambda}}_X)]^T$, and $\boldsymbol{r}_\varepsilon(\boldsymbol{x}_*) = [R_\varepsilon(\boldsymbol{x}_*, \boldsymbol{x}_1), \ldots, R_\varepsilon(\boldsymbol{x}_*, \boldsymbol{x}_m)]^T$.

The hyper-parameters include $\sigma_\varepsilon^2, \boldsymbol{\ell}$ and $\tau$ in the covariance function and are obtained by maximum likelihood estimation (MLE). By setting the derivative of equation (14) w.r.t. $\sigma_\varepsilon^2$ and solving, we find MLE of

$$\hat{\sigma}_\varepsilon^2 = \frac{(\boldsymbol{y} - \boldsymbol{H}\hat{\boldsymbol{\beta}})^T \boldsymbol{R}_\varepsilon^{-1} (\boldsymbol{y} - \boldsymbol{H}\hat{\boldsymbol{\beta}})}{m}. \tag{18}$$

Substituting equation (18) into (14) yields the concentrated log likelihood (Le Gratiet (2013))

$$-\frac{m}{2}\log(\hat{\sigma}_\ell^2) - \frac{1}{2}\log|\boldsymbol{R}_\varepsilon|, \tag{19}$$

and $\boldsymbol{\ell}$ and $\tau$ are found by maximizing this equation. Numerical method using a suitable global search routine such as genetic algorithm (GA) is needed for this optimization program. Nevertheless, a fully Bayesian analysis for finding those hyper-parameters can also be adopted (Bastos and O'Hagan (2009)).

## 4. EXPERIMENTAL DESIGN

The technique of *design of experiment* is an important link between experimental and modeling world. In an early stage of process modeling when a priori knowledge is rare, design methods which embody space-filling property are suggested. Kalagnanam and Diwekar (1997) proposed Hammersley sequence sampling (HSS) technique to generate data points that spread out "evenly" over the design region, and was later adopted by Yan et al. (2011). HSS is an efficient design method in many areas, because it requires fewer data compared to other space-filling methods, while they achieve similar performance. In light of its good property and easy implementation, we use HSS to generate design points for building base model and model migration. The HSS points can be generally obtained via the flowing steps (Kalagnanam and Diwekar (1997)):

(1) Any non-negative integer $n$ can be represented as a radix notation of a prime $p$ as follows:

$$n = n_m n_{m-1} \dots n_0 = n_0 + n_1 p + n_2 p^2 + \cdots + n_m p^m,$$

with integers $n_i \in [0, p-1]$, and $m$ being the integer part of $\log_p n$.

(2) The point in $[0,1]$ is obtained by reversing the order of the bits of $n$ and moving the decimal point:

$$\phi_p(n) = .n_0 n_1 \dots n_m = n_0 p^{-1} + n_1 p^{-2} + n_2 p^{-3} + \cdots + n_m p^{m-1}$$

(3) For the first $(d-1)$ primes $p_1, \dots, p_{d-1}$ , the Hammersley points on $d$-th dimensional space are given:

$$\boldsymbol{x}_k = \mathbf{1} - \left( \frac{k}{n}, \phi_{p_1}(k), \dots, \phi_{p_{d-1}}(k) \right), \quad k = 1, \dots, n,$$

with $\mathbf{1}$ being the $1 \times d$ unity vector. Note that $\boldsymbol{x}_k \in [0,1]$ for $k = 1, \dots, n$.

Abundant data points are required for accurately developing a base (GP) model. In this paper, a rule of thumb for the number points that should be used in the experimental design of old process is $n = 10d$ (Loeppky et al. (2009)). Sometimes $n$ can be greater than this, allowing us to build a more accurate base model, and if the relationship between base and new model is simple, the required data points for migration, $n$, may be somewhat fewer. We will return to this subject in Section 5.

## 5. EXAMPLE

A continuously stirred tank reactor (CSTR, see Kalagnanam and Diwekar (1997) and Yan et al. (2011) for example) is investigated to demonstrate the benefits of the proposed method. The process selected in this study consists of a first-order sequential reaction, A $\rightarrow$ B $\rightarrow$ C, taking place in a nonisothermal CSTR. The response of interest is the production rate of species B ($R_B$), which is mainly affected by five input variables: inlet concentration of A ($C_{Ai}$), inlet concentration of B ($C_{Bi}$), reaction temperature ($T$), the reactor volume ($V$), and volumetric flow rate ($F$). The production of species B and the steady-state values of other variables in the CSTR are governed by the equations (20)-(24). The average residence time ($\tau$) of each species in the reactor is given $\tau = V/F$:

$$C_A = \frac{C_{Ai}}{1 + k_A^0 \exp(-E_A/RT)\tau} \tag{20}$$

$$C_B = \frac{C_{Bi} + k_A^0 \exp(-E_A/RT) C_A}{k_B^0 \exp(-E_B/RT)\tau} \tag{21}$$

$$-r_A = k_A^0 \exp(-E_A/RT) C_A \tag{22}$$

$$-r_B = k_B^0 \exp(-E_B/RT) C_B - k_A^0 \exp(-E_A/RT) C_A \tag{23}$$

and

$$R_B = r_B V. \tag{24}$$

According to the literatures, two reactions varying from kinetic parameters $k^0$ and $E$ are selected to simulate similar, but non-identical processes. Their parameters and input ranges are summarized in Table 1.

We will now build base model using GP technique as discussed in Section 2. For old process, 50 HSS points are designed to produce the response variables of species B. In addition, a Gaussian random noise with variance $10^{-4}$ is added to the responses to simulate measurement errors. The data are beforehand normalized to satisfy the conditions with mean zero and variance one, and the predictive performance of the final model is assessed by the root-mean-square-errors (RMSE):

$$\sqrt{\frac{\sum_{i=1}^{n_t} [\hat{y}(\boldsymbol{x}_i) - y(\boldsymbol{x}_i)]^2}{n_t}}, \tag{25}$$

where $\hat{y}(\boldsymbol{x}_i)$ is the prediction and $y(\boldsymbol{x}_i)$ is the real response from experiments; $n_t$ is the number of testing points. The RMSE for base model evaluated at 200 HSS-generated testing points is 9.7560.

We will now exam whether the proposed method really has the potential to enhance model performance as stated in the introductory section. Similarly to base model, 200 HSS points are generated from new process for testing purpose. Given the fact that varying kinetic parameters in the reactions does not change governing mechanisms for both processes (therefore the process difference is simple), we set $\rho_{X,k} = 1$ for $k = 1, \dots, 5$. In the first stage of parameter estimation, the bias parameters $\boldsymbol{\lambda}_X$ will be found by minimizing the sum of squared prediction errors as stated in (8). The resulting $\hat{\boldsymbol{\lambda}}_X$ is then substituted into (12) in the second stage estimation, and the predictions of new model will be given as posterior means through Bayesian adjustments. Different numbers of training data are considered (the generation of training data strictly follows HSS technique). Figure 1 shows the results of RMSE w.r.t. to the number of training data, where "Base" refers to the predictions made from base model, "GP" means the new GP model developed from new process data only, "SBC" refers to slope/bias correction method proposed by Lu and Gao (2008b), and "Proposed" is our method using Bayesian adjustments. From the figure, it is clear that the base model is not ideally suitable to describing new process because of the model discrepancy, and we also note the classical GP model is not capable of accurately modeling the new process when there are only a few training data. SBC method has shown improved performance when the training data are limited; however, the RMSE does not reduce as expected when training data grow, i.e., the RMSE values seem nearly stable after adding 30 or more data. The benefits of the proposed method are evident from the figure: first, it has equal performance as compared to SBC if fewer data are available; second, further RMSE reduction can be guaranteed if there are sufficiently large amount of data – the advantage of the Bayesian adjustments method. Furthermore, the behaviors of GP and the proposed methods begin to converge when abundant training data are available, a satisfactory performance of our method.

Next we focus on a typical number of training data to further illustrate our method, for instance, 20 data points are gener-

Table 1. Parameters and their values for old and new processes

| Parameters | Values | | Units | Description |
|---|---|---|---|---|
| | Old process | New process | | |
| $k_A^0$ | $8.4 \times 10^5$ | $1.0 \times 10^6$ | $\mathrm{min}^{-1}$ | Physical constant |
| $k_B^0$ | $7.6 \times 10^4$ | $6.0 \times 10^4$ | $\mathrm{min}^{-1}$ | Physical constant |
| $E_A$ | $3.64 \times 10^4$ | $3.04 \times 10^4$ | J/mol | Physical constant |
| $E_B$ | $3.46 \times 10^4$ | $3.66 \times 10^4$ | J/mol | Physical constant |
| $R$ | 8.314 | 8.314 | $\mathrm{J/(mol \cdot K)}$ | Physical constant |
| $C_{Ai}$ | 1000-5000 | 800-4000 | $\mathrm{mol/min^3}$ | Input variable |
| $C_{Bi}$ | 100-500 | 100-500 | $\mathrm{mol/min^3}$ | Input variable |
| $T$ | 290-330 | 290-380 | K | Input variable |
| $V$ | .01-.09 | .01-.09 | $\mathrm{m^3}$ | Input variable |
| $F$ | .01-.09 | .01-.09 | $\mathrm{m^3/min}$ | Input variable |

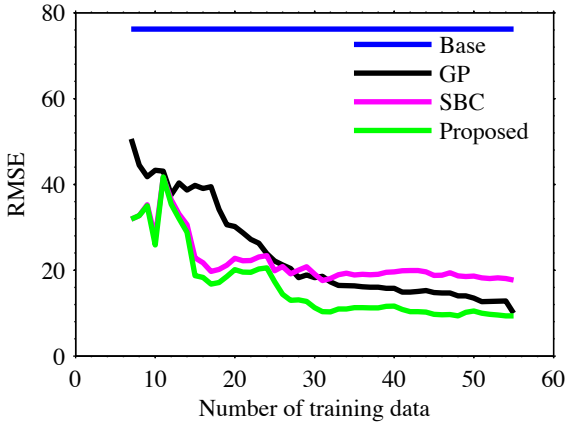


Fig. 1. Comparison of the different methods with respect to predictive RMSE

ated from the new process, a fraction of the data requirement for the base model training. The parameters characterizing the Bayesian adjustments are summarized in Table 2. Figure 2 shows the plot between the predictions using Bayesian adjustments and the real values, re-confirming the superiority of the proposed methods over base model prediction, GP and SBC methods. It is also worth noting that the Bayesian adjustments attains similar RMSE compared to GP method (14.3596 versus 14.6688), but requires a fraction cost of that method (20 versus 41), a desirable behavior of model migration.

Table 2. Parameters for the Bayesian adjustments

| | | | | | |
|---|---|---|---|---|---|
| $\hat{\boldsymbol{\lambda}}_Y$ | $\hat{\lambda}_{Y,1}$ | -0.2804 | $\hat{\boldsymbol{\beta}}$ | $\rho_Y$ | 0.4109 |
| | $\hat{\lambda}_{Y,2}$ | 0.0817 | | $\hat{\lambda}_0$ | 0.1140 |
| | $\hat{\lambda}_{Y,3}$ | 0.6568 | | $\hat{\lambda}_1$ | 0.0391 |
| | $\hat{\lambda}_{Y,4}$ | -0.1703 | | $\hat{\lambda}_2$ | -0.0184 |
| | $\hat{\lambda}_{Y,5}$ | 0.0494 | | $\hat{\lambda}_3$ | 0.1794 |
| $\hat{\boldsymbol{\ell}}$ | $\hat{\ell}_1$ | 0.1132 | | $\hat{\lambda}_4$ | 0.0021 |
| | $\hat{\ell}_2$ | 0.0116 | | $\hat{\lambda}_5$ | 1.0812 |
| | $\hat{\ell}_3$ | 0.0043 | $\hat{\boldsymbol{\sigma}}^2$ | $\hat{\sigma}^2_{\ell}$ | 4.5603 |
| | $\hat{\ell}_4$ | 0.0128 | | $\hat{\sigma}^2_{\varepsilon}$ | 0.0161 |
| | $\hat{\ell}_5$ | 0.5000 | | | |

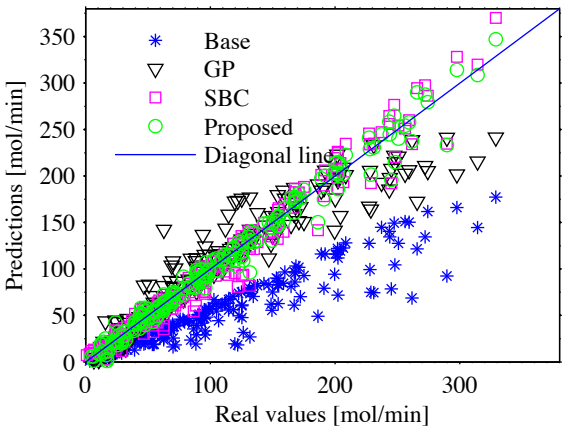


Fig. 2. Comparison of the different methods with respect to prediction versus real values plot

## 6. CONCLUSION

In this paper, a Bayesian adjustments strategy is proposed to enhance model migration, and a two-stage estimation procedure is suggested to ensure the strategy's feasibility and obtain a close-form solution. Our strategy is common in, but not limited to, similar, yet identical processes of which the differences are simple and systematic. The efficiency of the strategy is demonstrated on a simulated chemical reaction.

Several directions can be extended for future research. For example, Bayesian adjustments model migration must be closely examined under conditions in which the base model is not accurate. Second, the Bayesian adjustments strategy can be applied to process optimization. A common optimization problem may seek for the maximum or minimum of a model-oriented objection function. Efficiency, in terms of the number of training data that should be generated from the process, is expected by adopting Bayesian adjustments migration.

## REFERENCES

Bastos, L.S. and O'Hagan, A. (2009). Diagnostics for gaussian process emulators. *Technometrics*, 51(4), 425–438.

Fearn, T. (2001). Review: Standardisation and calibration transfer for near infrared instruments: a review. *Journal of Near Infrared Spectroscopy*, 9(4), 229–244.

Feudale, R.N., Woody, N.A., Tan, H., Myles, A.J., Brown, S.D., and Ferré, J. (2002). Transfer of multivariate calibration models: a review. *Chemometrics and Intelligent Laboratory Systems*, 64(2), 181–192.

Kalagnanam, J.R. and Diwekar, U.M. (1997). An efficient sampling technique for off-line quality control. *Technometrics*, 39(3), 308–319.

Le Gratiet, L. (2013). Bayesian analysis of hierarchical multi-fidelity codes. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1), 244–269.

Loeppky, J.L., Sacks, J., and Welch, W.J. (2009). Choosing the sample size of a computer experiment: A practical guide. *Technometrics*, 51(4), 366–376.

Lu, J. and Gao, F. (2008a). Model migration with inclusive similarity for development of a new process model. *Industrial & Engineering Chemistry Research*, 47(23), 9508–9516.

Lu, J. and Gao, F. (2008b). Process modeling based on process similarity. *Industrial & Engineering Chemistry Research*, 47(6), 1967–1974.

Lu, J., Yao, K., and Gao, F. (2009). Process similarity and developing new process models through migration. *AIChE journal*, 55(9), 2318–2328.

Lu, J., Yao, Y., and Gao, F. (2008). Model migration for development of a new process model. *Industrial & Engineering Chemistry Research*, 48(21), 9603–9610.

Qian, P.Z. and Wu, C.J. (2008). Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics*, 50(2), 192–204.

Rasmussen, C.E. (2006). *Gaussian processes for machine learning*. MIT Press.

Vastola, J.T., Lu, J.c., Casciato, M.J., Hess, D.W., and Grove, M.A. (2013). A framework for initial experimental design in the presence of competing prior knowledge. *Journal of Quality Technology*, (4), 301–329.

Yan, W., Hu, S., Yang, Y., Gao, F., and Chen, T. (2011). Bayesian migration of gaussian process regression for rapid process modeling and optimization. *Chemical Engineering Journal*, 166(3), 1095 – 1103.