# Model Migration for Development of a New Process Model

**ARTICLE** *in* INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH · DECEMBER 2008

Impact Factor: 2.59 · DOI: 10.1021/ie8013296



**3 AUTHORS:**

Junde Lu
The Hong Kong University of Science and T…
**6** PUBLICATIONS **66** CITATIONS
SEE PROFILE

Yuan Yao
National Tsing Hua University
**52** PUBLICATIONS **413** CITATIONS
SEE PROFILE

Furong Gao
The Hong Kong University of Science and T…
**267** PUBLICATIONS **3,781** CITATIONS
SEE PROFILE

# Model Migration for Development of a New Process Model

**Junde Lu, Yuan Yao, and Furong Gao***

*Department of Chemical Engineering, The Hong Kong University of Science & Technology, Clearwater Bay, Kowloon, Hong Kong*

Data-based process models are usually developed by fitting input−output data collected on a particular process. The model built on one particular process becomes invalid with another similar process. Traditional data-based modeling methods have to completely rebuild a new process model on a similar process, leading to repetition of a large number of experiments, if process similarities between two similar processes are ignored. Effective use and extraction of these process similarities and migration of the existing process model to the new process can require a fewer number of experiments for the development of a new process model, resulting in savings of time, cost, and effort. In this paper, we present a model migration method that can quickly model a new process based on an existing base model and contrast information between the base model and the new process. The method developed involves a procedure of six steps: information extraction from the base model, initial design of experiments, slope/bias correction (SBC) to the base model, outlier detection and assessment, further design of experiments, and development of the new model by combining local difference models and the corrected base model. An example is provided to illustrate the new model development strategy for predicting injection molded part weight, taking advantage of an existing model.

## 1. Introduction

Development of a process model involves finding a relationship between a set of X variables (process conditions or inputs) and one or several response variables Y (quality properties, outputs). Generally, process models can be obtained by using first-principle, empirical, or hybrid approaches. First-principle models, based on prior process knowledge, represent process behavior through a set of conservation equations. They allow for a good degree of extrapolation. It is, however, difficult and time-consuming to develop such a model with sufficient accuracy because of the mismatch of process conditions and simplifications that are inevitable. Solutions of these models are also computation heavy. On the other hand, empirical approaches, such as artificial neural networks (ANNs), fuzzy logic model (FLMs), partial least-squares (PLS), and support vector machines (SVMs), use a set of experimental input−output data to be fitted as the relationship between process input and product quality.[1−5] They allow only limited extrapolation beyond the domain of training data (extrapolation). To obtain good extrapolation properties with sufficient accuracy, hybrid models, which combine prior knowledge or first-principle and traditional data-driven methods, have been proposed for the chemical process.[6,7]

Industrial batch processes often produce products of different grades by changing operating conditions and/or feedstock properties, to reflect changes in quality specifications. Considering two similar processes, $P_a$ and $P_b$, although they make different products under different process conditions, their intrinsic physical principles are the similar. The process model developed on process $P_a$ might not be directly applicable to process $P_b$. With data-based modeling, this problem necessitates repetition of the experiments, remeasurement of samples, and redevelopment of a new model for the new process. It thus requires considerable cost, effort, and a large time investment. For clarity, we define the existing model describing process A as a base model and a model to be developed for a new process as the new model.

* Corresponding author. E-mail: kefgao@ust.hk. Tel: +852-2358-7139. Fax: +852-2358-0054.

As we know, despite operating conditions that may be different for different batch processes, certain process behaviors and characteristics remain common under these different conditions. For example, in injection molding, where plastic granules are processed into different specifications of product with different molds, process conditions, such as barrel temperature, packing pressure, and injection velocity, have similar impacts on the molded part qualities. This common characteristic, to a degree, should have been already built in the data-driven model, i.e., the base model. This motivated us to develop a method that allows the model to migrate from the old process to a new one. If the model (base model) of the old process is available, how can we extract the information on the base model so to design fewer but efficient experiments to describe the features of new process? On the basis of new collected data from the new process, how can we migrate the base model to develop a model for the new process?

In this paper, we utilize the concept of model migration to develop a model for a new process, therefore building a new model using less experimental data to save time and cost (see Figure 1). In developing this new model, common process behaviors and characteristics, or process similarity, are explored by taking advantage of the existing base model. The method developed can be applied to either batch processes or continuous processes. The requirements and challenges of the concept have been stated by Lu and Gao.[8] A general framework for model migration and a simple case where the process difference between two similar processes is a shift and rescale problem have been given as well. However, this paper is limited to the

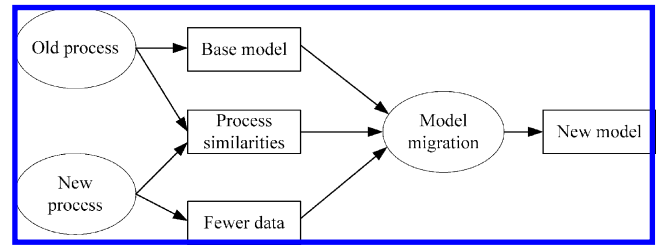


**Figure 1.** Schematic of model migration strategy.

10.1021/ie8013296 CCC: $40.75 
Published on Web 12/02/2008

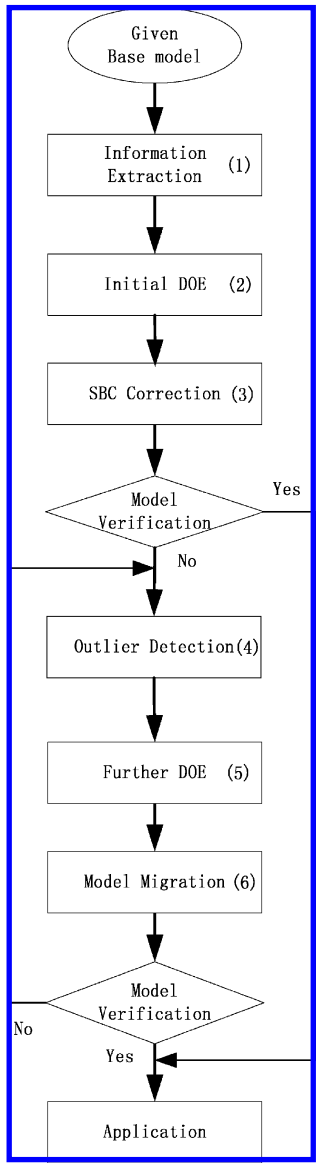


**Figure 2.** Block diagram for developing a new model based on the model migration strategy.

case where the process difference is merely shift and rescale. This paper further develops the concept of model migration for the case where the process similarity is unknown and needs exploration.

The remainder of the paper is organized as follows. The model migration strategies are presented in section 2. Section 3 gives an illustrative application on the injection molding process. Finally, some concluding remarks are given in section 4.

## 2. Model Migration Strategies

A recursive model migration procedure for the development of a new model is proposed, which includes the following few steps: (1) process information extraction from the base model, (2) initial design of experiment (IDOE), (3) slope and bias correction (SBC) to base model, (4) outlier detection, (5) further design of experiment (FDOE), and (6) model migration, as shown in Figure 2. The procedure begins with the given base model, followed by the initial DOE step guided by process information extracted from the base model. The objective of the initial DOE is to design some sparse experiments to determine if the base model is still valid or not. If not, a slope/bias (SBC) correction of the base model is therefore conducted to correct a shift and rescale the difference by using initial DOE data. In an industrial process, two similar processes often have shift and rescale differences in the input and output space, such as the process from laboratory-scale to pilot-scale. An SBC correction is a mathematic transformation of input and output variables between the base model and the new model. Such transformation includes a slope and bias transformation of input and output space. Thus, a test can be conducted to check whether the new process is a shift and rescale to the old process. If all data points can be fitted by a shift and rescale model, then the new process is simply a shift and rescale to the old process and thus the new model is simply the shift and rescale corrected base model. If not, there exist some data points which do not fit the corrected model well, and we define them as outliers. The outliers detected can assist further DOE to determine more details on the mismatch of the corrected model. Finally, a model migration strategy is conducted to develop the new model by using both initial and further DOE data based on the corrected base model. The model verification is then conducted to verify satisfactory performance of the new model. If the verification result is not satisfactory, the procedure will go back to step 4. A graphical representation of the procedure that realizes the proposed approach is shown in Figure 2. The key requirements, challenges, and possible solutions for each step are given as follows.

**2.1. Process Information Extraction from Base Model.** Despite development with input−output data alone, the base model, as pointed out in the Introduction, should contain certain key process characteristics. These process characteristics could be extracted to assist the understanding of the old process and to guide the experimental design for the new process.

To investigate the impact of each input variable (factor) on the response variable, the main effect of each factor is introduced. The main effect (ME) of factor A in a two-level design is the difference between the average response at the high (+) level of A and the average response at the low (−) level of A, as shown in eq 1.

$$\mathrm{ME(A)} = \bar{y}(\mathrm{A}+) - \bar{y}(\mathrm{A}-) \tag{1}$$

Generally, A+ and A− are used to represent the high and low levels of A, respectively.[9] Calculation of the main effect of each factor is based on input−output data generated from the base model. Factors with an significant main effect should be specified with more factor levels when designing initial experiments for the new process. New experiments designed by such a strategy will have good representation of the new process and can significantly improve the model prediction precision, as we can focus on the important factors and take less consideration for less important factors, to reduce the number of experiments.

**2.2. Initial Design of the Experiment.** Traditionally, training data for modeling is collected through statistical design of the experiment (DOE). Generally, without any process information, an experimental program begins with a screening design considering only two levels for each factor, followed by an augmented experiment to develop a more accurate model.[10] Geometrically, the conventional DOE approach is an even grid of experiments on a response surface spanned by input variables, which apply equivalent weights on each data. Obviously, this type of design may not be optimal for the new process with availability of the base model. Actually, because of the nonlinearity of the process input−output relationship, output may not be evenly distributed with evenly distributed inputs. For example, considering a one-dimensional example, Figure

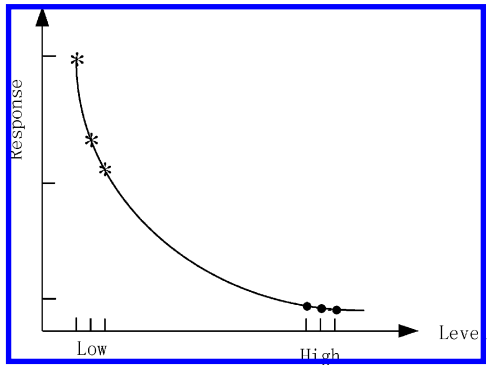


**Figure 3.** Unevenly distributed response.

3 shows the nonlinear relationship between input factor $x$ and the output response $y$ by

$$y = f(x) \tag{2}$$

Evenly distributed factor levels will result in an unevenly distributed y value through the mapping of function*f*. The factor at the low level region has significant impact on the response. In other words, when considering the input and output space together, the factor at the low level region where data points have low data density has a significant effect on response, such as the star data points shown in Figure 3. Therefore, in order to better describe such nonlinear change of response and improve model prediction ability, more experiments should be performed at regions with low data density.

With the availability of a base model, an efficient experimental design for the new process can be obtained by analyzing the nonlinear input−output relationship of the base model. This design may not be an optimal for the new process, but it must be a good starting design, as these two processes are assumed to have good similarity.

Good experimentation should be on a sequential or hierarchical basis, so that, in the beginning, some sparse experiments should be conducted to capture the key features of the new process. Therefore, the objective of initial DOE is to design some sparse but influential and critical experiments which can efficiently reflect the new process's behavior.

Cluster estimation, developed by Chiu, can group a large data set into several cluster centers, each cluster center being most representative of a local region.[11] Clustering is a density-based clustering method in spatial nature. Application of such spatial clustering to input−output space data generated by the base model can naturally produce cluster centers that can reflect density of data points. The lower the density of data points, the more cluster centers expected. Thus, a nonlinear input−output relationship of the base model is described by unevenly distributed cluster centers. For the clustering test, a data set can be readily generated at no experimental cost, as the base model can be used to describe the old process behavior at any input points. Generation of the data set ($N$ data points) starts with discretization of input−output space ($Z = [X^T;y]$) by griding $X$ and $y$ with equidistant lines on the base model, where $X$ denotes inputs, $y$ denotes output, and $Z$ is the augmented data vector, respectively. By applying the clustering method on the generated data set, obtained cluster centers are critical and influential data points which can efficiently reflect the old process behavior. This kind of experimental design on an old process can thus be migrated to a new process by conducting experiments at each center to capture the new process behavior.

The clustering algorithm considers each data point $Z_i$ as a candidate prototype cluster center. The capability of a point to be a cluster center is evaluated through its potential, a measure of the spatial proximity between current point $z_i$ and all other data points:

$$P_i = \sum_{j=1}^{N} \exp\left(-\frac{4}{r^2}\|z_i - z_j\|^2\right) \quad i \in [1, N] \tag{3}$$

where $P_i$ represents the potential of the $i$th data point, $N$ denotes the number of data points surrounding the potential center, and $r$ is effective radius of a cluster center, defining a neighborhood; data points outside this radius have little influence on the potential. Thus, the measure of potential for a candidate data point is a function of its distances to all other data points. Traditional subtractive clustering uses a uniform radius for each dimension of input−output space, which makes effective ranges of cluster centered on each dimension equal. In fact, based on the main effect calculated from the base model, input variables may have different impacts on output. To apply such different input impacts into initial DOE, a nonuniform effective radius based for measure of the potential of data point $Z_i$ is proposed:

$$P_i = \sum_{j=1}^{N} \exp\left\{-\left(\frac{4\|z_{i,1} - z_{j,1}\|^2}{r_1^2} + \frac{4\|z_{i,2} - z_{j,2}\|^2}{r_2^2} + \cdots \frac{4\|z_{i,m+1} - z_{j,m+1}\|^2}{r_{m+1}^2}\right)\right\} = \sum_{j=1}^{N} \exp\left(-\sum_{k=1}^{m+1} \frac{4\|z_{i,k} - z_{j,k}\|^2}{r_k^2}\right) \tag{4}$$

where $k = 1,...,m + 1$ is the dimension of input−output space, $m$ is the dimension of input space, and $m + 1$ represents the output space (for only one output case). $Z_{i,k}$ and $Z_{j,k}$ represent the $k$th dimension of data points $i$ and $j$, respectively. Selection of an effective radii vector is based on the main effect of the corresponding input variables; the larger the main effect, the smaller the radius. Thus, the influential space of each center is changed from hypersphere to hyperellipsoid. Detailed literature on finding the cluster centers can be found in Chiu's report.[11]

By properly setting the effective radius of each dimension, the number and location of cluster centers can be found. The region where input variables have significant impact on output will have a lower density of data points, and more cluster centers are expected to be found, based on the clustering algorithm and evenly distributed data generated from the base model. Each cluster center is in essence a prototypical data point that exemplifies a characteristic behavior of the old process. Naturally, new experiments conducted on such cluster centers may potentially describe the key features of the new process. By applying the main effect of a factor into the clustering algorithm, nonlinear impacts of the factor are examined.

**2.3. Slope/Bias Correction (SBC).** After collection of initial DOE data for a new process, a simply test is conducted to test if the base model can be directly used. If not, a slope/bias correction is conducted to correct the shift and rescale the difference. Generally, differences between two similar processes vary from the simplest case, where the new process is simply shift and rescale of the old process, to the other case, where the process similarity is unknown. For the simplest case, an output slope/bias correction (OSBC) or input−output slope/bias correction (IOSBC) can be applied to the base model.

The slope/bias correction generally involves input and output correction. Given the base model of any form:

$$Y_{\text{base}} = f(X_{\text{base}}) \tag{5}$$

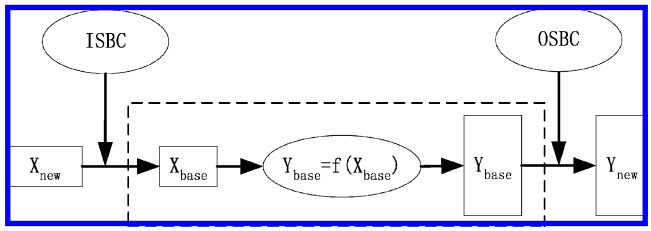


**Figure 4.** Slope and bias correction.

where $X_{base}$ and $Y_{base}$ are the input and output of base model, respectively, and $f$ is any nonlinear function to describe the old process. If there is only a shift and rescale in output space, we can expect that there is a linear relationship between predicted output $Y_{base}$ from base model and measured output $Y_{new}$ from new process when using the same input variables.

$$Y_{new} = S_O \times Y_{base} + B_O = S_O \times f(X_{base}) + B_O \quad (6)$$

where $S_O$ and $B_O$ represent rescale and shift parameters of the old process in output space, respectively. This correction of predicted output is called the output slope/bias correction (OSBC), as illustrated in Figure 4. Estimation of the above parameters can be obtained through ordinary least-squares using data collected from the initial DOE on the new process.

OSBC can only be applied on the case where two similar processes have the same operating range. More generally, because of different production scales or other cases, two similar processes might have different operating ranges. If we merely apply OSBC on this situation, significant error will be found. A solution to this is to apply an input slope/bias correction (ISBC) on input space to correct the operating range difference. In addition, complicated differences may lead to shift and rescale in the input space for the base model. Therefore, for these two cases, the inputs $X_{new}$ measured for the new process should be transformed to those of the base model ($X_{base}$), also called input slope/bias correction (ISBC), as eq 7 shows:

$$X_{base} = S_I \times X_{new} + B_I \quad (7)$$

where $S_I$ and $B_I$ denote rescale and shift parameters of the old process in input space, respectively. Thus, a new model is obtained by combining an input slope/bias correction (ISBC) and an output slope/bias correction (OSBC) of the base model (see Figure 4):

$$Y_{new} = S_O \times f(S_I \times X_{new} + B_I) + B_O \quad (8)$$

This kind of correction is called input−output slope/bias correction (IOSBC). Independent of whether the base model structure is known or not, estimation of the above parameters can be obtained by optimizing the following equations together with the initial DOE data from the new process:

$$\begin{cases} \arg\min J(S_O, B_O, S_I, B_I) = \varepsilon\varepsilon^T S_O, B_O, S_I, B_I \\ \varepsilon = y - \{S_0 \times f(S_1 \times X_{new} + B_1) + B_0\} \end{cases} \quad (9)$$

where $y$ is observed data on the new process and $\varepsilon$ denotes the predicted errors between the observed values on new process and predicted values from the corrected base model. All identified shift and rescale parameters are vectors determined by the dimension of input and output space.

When we apply IOSBC on the base model, not all factors or inputs are transformed. Only the factors with a significant main effect are considered, which can reduce the new experiments needed to develop the new model, as the number of new experiments depends on the number of parameters to be identified in IOSBC. In general, OSBC or IOSBC will work well when the differences between the two similar processes are simple. Since only several parameters are needed to be identified, time and effort needed to build such a new model is small. However, when the process differences are more complex, such a strategy may not obtain satisfactory results. A comprehensive model migration algorithm stated later is needed to obtain acceptable prediction results.

**2.4. Outlier Detection and Assessment.** In multiple linear regression model building, the data used for parameter estimation sometimes violates the assumptions of least-squares. These "bad" data are designated as regression outliers because they deviate from the linear relation followed by the bulk of the data.[12] The same problem arises when a slope/bias correction is applied to the base model. Since the new process may not be merely a shift and rescale to the base model, there are some data on which relatively larger predicted errors are found on the corrected base model. Such data contain valuable information for improving prediction of the corrected base model. We define such data as shift/scale outliers or SS outliers.

For the outlier detection, all new data are regarded as candidate SS outliers. Often, the model prediction error is of primary interest to users of regression models, so SS outliers can be identified by examination of change in the fit if the $i$th candidate SS outlier is deleted. To assess the sensitivity of the change in prediction of the corrected base model to the deletion of the $i$th candidate SS outlier, we compare the prediction ability of the corrected base model that is corrected by $n$ candidate SS outliers with that corrected by the remaining $n - 1$ candidate SS outliers, that is

$$\lambda(i) = |\sum_{j=1}^{n} (y_j - \hat{y}_j) - \sum_{j=1}^{n} (y_j - \hat{y}_j(i))| \quad (10)$$

where $\lambda(i)$ denotes deviation index or change in prediction when $i$th candidate SS outlier is deleted, n represents number of candidate SS outliers which is used to correct the base model, $y_j$ is actual value of $j$th candidate SS outlier, and $\hat{y}_j$ and $\hat{y}_j(i)$ represent predicted values of the corrected base model corrected by $n$ candidate outliers and that corrected by remaining $n - 1$ candidate outliers (the $i$th candidate outlier is ignored), respectively. Obviously, a threshold for determination of whether a candidate SS outlier is a true outlier has to be defined. A large threshold generally results in fewer SS outliers, while a small threshold can produce an excessive number of outliers, resulting in great effort to conduct a possibly excessive number of new experiments.

**2.5. Further Design of the Experiment.** Further DOE is to compensate for the insufficiency of initial DOE for model migration. Cluster estimation is still used for further DOE. Similar to initial DOE, a number of new cluster centers are needed to represent further DOE data points. This involves recomputing the potential value of each point and choosing the data point with the highest potential value as a new clustering center. Computation of potential values should consider changed effective radii, effect of outlier, and effect of the existing cluster center of initial DOE together.

**2.5.1. Changed Effective Radii.** Further DOE is to further cluster data points at a higher level to find more new cluster centers except for previously existing cluster centers, similar to the divisive clustering method. If the same effective radii $r$ for initial DOE are used, newly found cluster centers may overlap with the previously existing cluster centers. Therefore, smaller effective radii $r$ which define influential radii are used for recomputing the potential value of each point. Revised

calculation of the potential value of each data point with a smaller effective radii $r_F$ is:

$$P_i = \sum_{j=1}^{N} \exp\left\{-\left(\frac{4\|z_{i,1}-z_{j,1}\|^2}{r_{F,1}^2} + \frac{4\|z_{i,2}-z_{j,2}\|^2}{r_{F,2}^2} + \cdots \frac{4\|z_{i,m+1}-z_{j,m+1}\|^2}{r_{F,m+1}^2}\right)\right\} = \sum_{j=1}^{N} \exp\left(-\sum_{k=1}^{m+1} \frac{4\|z_{i,k}-z_{j,k}\|^2}{r_{F,k}^2}\right) \quad (11)$$

where $r_F$ represents new effective radii which are smaller than those used in initial DOE. A trial and error method can be used to determine the effective radii, as initial DOE is conducted on an existing base model without significant effort except for computation.

**2.5.2. Effect of Outlier.** Since outliers contain model mismatch information which can potentially improve new model prediction ability, new cluster centers should be as close as possible to the outliers. Such outlier impact should and can be considered by adding its potential to each data point:

$$P_i \leftarrow P_i + \sum_{m=1}^{M} \lambda_m P_m e^{\frac{-4\|z_i - z_m\|^2}{r_F^2}} \quad m \in [1, M] \quad (12)$$

where $M$ denotes the number of outlier points, $\lambda_m$ the deviation index which describes how the $m$th outlier point deviated from the corrected base model from eq 10, and $P_m$ the revised potential value of the outlier calculated from eq 11. According to eq 12, the potential of each point is increased by an amount proportional to deviation index and the potential of the $m$th outlier and inversely proportional to the distance to it. That is, the data points near the outlier will have greatly increased potential. Therefore, data points around the outlier are likely to be selected as the new clustering center and to avoid leaking important information. By doing so, the impact of influential points is incorporated into the revised potential value of each data point.

**2.5.3. Effect of Existing Cluster Center.** New cluster centers should not be close to previously existing ones, as it is not reasonable to conduct two or more experimental runs around one point. Therefore, to avoid new cluster centers that are too close to existing ones, the potential value of each new data point is revised by subtracting the potential value of existing clustering centers, as eq 13 shows:

$$P_i \leftarrow P_i - \sum_{k=1}^{K} P_k e^{\frac{-4\|z_i - z_k\|^2}{r_F^2}} \quad k \in [1, K] \quad (13)$$

where $P_k$ denotes the revised potential value of $k$th existing centers and $K$ denotes the number of existing clustering centers. According to eq 13, the potential of each point is reduced by an amount proportional to the revised potential of the existing center and inversely proportional to the distance to it. That is, the data points near the existing clustering center will have greatly reduced potential. In the extreme case, the potential value of the data point which coincides with the existing cluster center (index $i$ is equal to $k$) is reduced to zero. Therefore, data points around the previously existing center are unlikely to be selected as the new cluster center and to avoid cluster centers too closely spaced or overlapped, consequently reducing experimental runs.

When the potential of all data points have been revised through eqs 11 to (13, the data point with the highest potential value is chosen as a new clustering center. The following algorithm concerning how to find out other new cluster centers is the same as that in initial DOE.

**2.6. Model Migration.** Depending on whether the structure of the base model is known to the user, there are two possibilities for model migration. One is that the base model is completely open to the user and the base model structure is known. In this case, certain model evolution strategies such as neural networks or the fuzzy method can be used to develop a new model.[13,14] The more general case is where a black-box-based model (for example, provided with a commercial software package) is given to the user, so that both the base model structure and parameters are unknown. This black-box model maps only the input−output relationship. This is the most challenging case and is the one upon which we focus in this paper. The model migration strategy should retain characteristics that are common to the old process, remove information that are no longer valid for the new process, and add, modify, or update the new information evolved in the new process.

Because of the complexity of differences between the base model and new process, there are still mismatches between the corrected base model and the new process. This discrepancy is reflected by DOE data (initial DOE and further DOE data points), as we have considered the outlier impact when designing further DOE. Therefore, by exploring the DOE data distribution, we can have a good understanding of process difference. Generally, the region with the larger amount of DOE data would indicate the larger process difference.

Because of the complexity of differences, it is difficult to describe the process difference using one difference model. A solution to this problem is using several local difference models to describe the process difference. These difference models are combined together to form a global difference model based on a certain strategy, as stated later. The new model can thus be constructed by adding the global difference model to the corrected base model by linear superposition, as shown in Figure 5. The development of the new model within this framework typically consists of the following tasks:

1. Decompose the process into several local regions.

2. Select the local difference model structure for each local region.

3. Find a strategy for combining the local difference models together.

**2.6.1. Local Regions.** The local regions can be obtained by applying clustering estimation on DOE data. The approach is based on an assumption that the DOE data are representative samples of the process difference. If the obtained clusters are projected onto the corresponding fuzzy sets, the local regions can be described by a set of antecedent parts of fuzzy rules. The total input region is thus partitioned into several local regions. Notice that the local regions are not really regions with hard boundaries but rather regions with soft boundaries. This means that there will be a gradual transition between local difference models when moving between local regions. This is implemented as an interpolation technique, which can also be viewed as a fuzzy inference system.[15] It is important to select an effective radius when clustering DOE data, as the number and size of the local difference region are highly dependent on the number of cluster centers. Obviously, there is a tradeoff between the number and size of local regions and the complexities of the local difference models.

**2.6.2. Local Difference Model Structure.** Generally, the complexity of the difference between the corrected base model and the new process is typically not uniform. Sometimes a simple low-order local difference model may be sufficient in

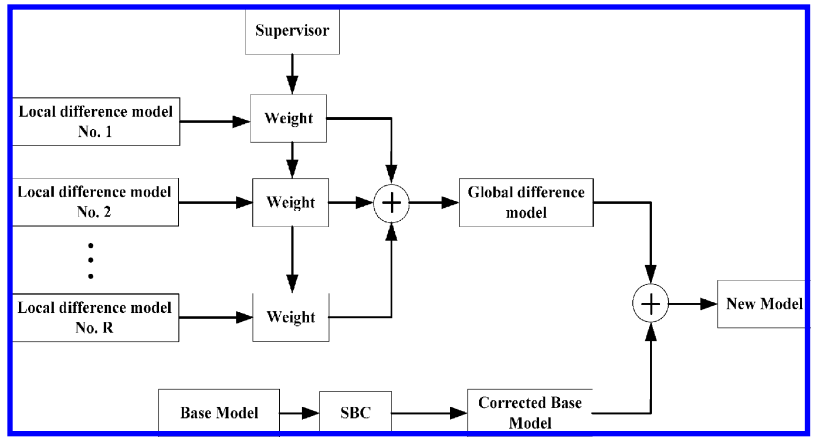


**Figure 5.** Development of a new model by bringing the corrected base model and difference models together.

some local regions, while in other cases a more complex local model has to be necessary in other local regions. Determination of each local model structure is one of the most challenging tasks in the development of a new model. The local model structures should be as robust and interpretable as possible, and the simplest one which can represent the local region to the desirable accuracy. The algorithms for structure identification should consider the complexity of the process, the representational ability of the local difference models, and the availability of DOE data. Note that the choice of size of local regions and local difference model structures should be coupled together.

**2.6.3. Combining Local Difference Models.** Assume that $R$ local difference models corresponding to local regions are given in the following form:

$$y_i = f_i(X = [x_1, x_2, ...x_n]); \quad i = \{1, R\} \tag{14}$$

where $y_i$ is output of $i$thlocal difference model, $R$ represents the number of local regions or rules, $X$ is the input vector, and $n$ is the dimension of input space. The global difference model is constructed in the same way as the TS model.[15] Therefore, description of each component of the global difference model is similar to the respective TS model component. The output of the global difference model is the weighted average of the $R$ local difference models:

$$y = \frac{\sum_{i=1}^{R} \varpi_i y_i}{\sum_{i=1}^{R} \varpi_i} \tag{15}$$

where $y$ is the output of the global difference model and $\varpi_i$ is the degree of membership to the $i$th local model.

Thereafter, the overall form of each local difference model is:

$$y_i\text{: if } (x_1 \text{ is } MF_{i1}) \text{ and } ... \text{ and } (x_n \text{ is } MF_{in}) \text{ then } (y_i = f_i(x_1, x_2, ... x_n)); \tag{16}$$

where $MF_{ij}$ are the fuzzy sets used to describe corresponding local difference regions. For the identification of the parameters of the local difference model, the weighted recursive least-squares estimation could be used, if the local difference models have linear structure.[13] For the nonlinear structure of the local model, optimization method has to be used to identify these parameters. Data used to identify the parameters is the difference between DOE data of the new process and predictions from the corrected base model. Notice that the number of DOE data in certain local regions should be sufficient to guarantee locally adequate behavior of the corresponding local model that forms the global difference model.

## 3. Case Studies

The injection molding, an important polymer processing technique, is an ideal batch process for the application and verification of the proposed methodology. It makes different shapes and types of products with different polymer materials and molds. The qualities of the product are the results of different combinations of material property, mold, and part geometry and processing conditions. Part weight, an important quality characteristic of the injection molding product, is selected as the product quality in this paper.

The experiments were performed on an 88 ton Chen Hsong reciprocating-screw injection molding machine, model number JM88MKIII. The process with mold type I and material of polypropylene (PP) is considered as the old process. Rather than only using different molds for the new process, the process with mold type II and material of high-density polyethylene (HDPE) is chosen as the new process. Thus, the new process is more than shift and rescale of the old process. The two mold geometries are shown in Figure 6 and Figure 7. From the screening design of experiment analysis, three process variables, including barrel temperature (BT), packing pressure (PP), and injection velocity (IV) are chosen to describe part weight, as they have relative significant effect on it. Then, a full factorial design with 27 experimental runs is conducted, each factor with three levels, to build the base model, as shown in Table 1. Through Minitab software, an analysis of variance, ANOVA, is made on experimental data and a quadratic polynomial regression model is built as the base model with the relative root-mean-square error (R-RMSE) of 0.415%. R-RMSE is defined as

$$R - \mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} \left( \frac{Y_i - \hat{Y}_i}{Y_i} \right)^2}{N}} \tag{17}$$

where $Y_i$ is the observed data from old process, $\hat{Y}_i$ is the estimated value from the base model, and $N$ is the number of data. Assuming the base model structure is unknown, with the exception of the input−output relationship, a set of data are generated from the base model to calculate the main effect of each factor according to the equation, and the results are shown in eq 20.

$$[\mathrm{BT}, \mathrm{PP}, \mathrm{IV}] = [-1.058, 4.288, -0.222] \tag{18}$$

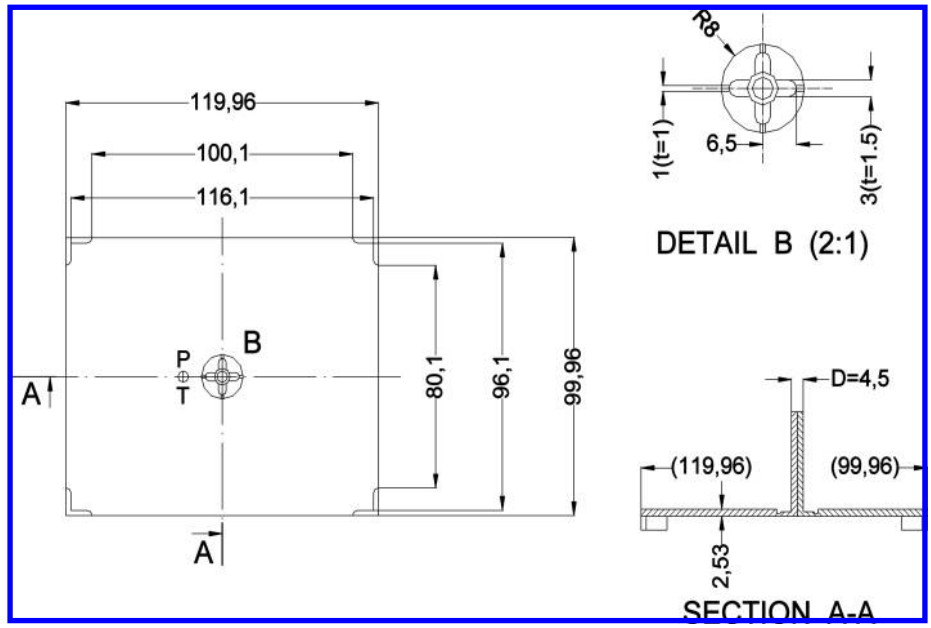


**Figure 6.** Geometry of mold type I with a flat cavity.

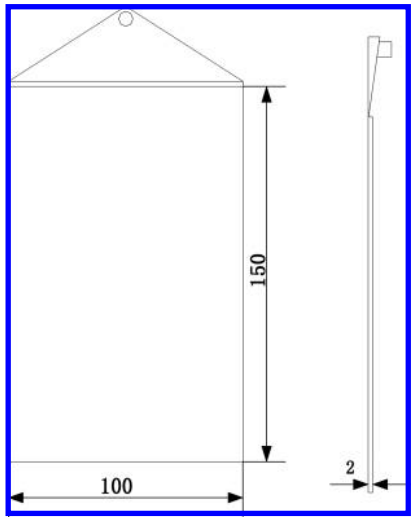


**Figure 7.** Geometry of mold type II with a fan gate.

**Table 1. Factors and Levels for Full Factorial Design**

| | level | | |
|---|---|---|---|
| factor | 1 | 2 | 3 |
| barrel temperature (BT)(°C) | 210 | 220 | 230 |
| packing pressure (PP) (bar) | 150 | 225 | 300 |
| injection velocity (IV) (mm/s) | 15 | 22.5 | 30 |

Calculated results show that packing pressure has significant positive impact on the part quality, and barrel temperature and injection velocity have relatively less significant negative impact on the part quality. Consequently, the three effective radii for clustering estimation are selected:

$$r = [r_1, r_2, r_3] = [0.75, 0.35, 0.9] \tag{19}$$

Then, a set of generated data points calculated from the base model are clustered to represent experimental conditions for the new process. Clustering results indicate that there are nine centers to represent the major features of the base model. Since the base model is built with data normalized to [0, 1], the clustered centers cannot be directly used on new process before transfer to new physical conditions, as new process variables may have different operating ranges. On the basis of machine specification and changes of mold and material, new ranges of each process variable are determined: (barrel temperature (BT): [200 220]; packing pressure (PP): [350 500]; injection velocity (IV): [24 40]). Although the settings of the new operating range may not be in accordance with the actual physical conditions, this does not have significant impact on model migration algorithm, as the successive input slope/bias correction (ISBC) will compensate for this discrepancy. Clustering results also indicate that barrel temperature, packing pressure, and injection velocity have six, seven, and five levels, respectively. This is consistent to previous analysis results that packing pressure has a more significant main effect than the other two, as shown in Table 2.

**Table 2. Cluster Centers and New Experiments on New Process**

| | BT (°C) | | PP (bar) | | IV (mm/s) | |
|---|---|---|---|---|---|---|
| experiments | center | transferred | center | transferred | center | transferred |
| 1 | 0.5 | 210 | 0.3 | 395 | 0.5 | 32.5 |
| 2 | 0.5 | 210 | 0.65 | 447.5 | 0.5 | 32.5 |
| 3 | 0.5 | 210 | 0.9 | 485 | 0.5 | 32.5 |
| 4 | 0.75 | 215 | 0.1 | 365 | 0.6 | 34 |
| 5 | 0.9 | 218 | 0.5 | 425 | 0.6 | 34 |
| 6 | 0.15 | 203 | 0.1 | 365 | 0.55 | 33.25 |
| 7 | 0.15 | 203 | 0.5 | 425 | 0.75 | 36.25 |
| 8 | 0.95 | 219 | 0.85 | 477.5 | 0.75 | 36.25 |
| 9 | 0.05 | 201 | 0.8 | 470 | 0.9 | 38.5 |

Having transferred the nine centers to corresponding physical conditions, these new experiments were performed on the new process, and an output slope/bias correction is applied on the base model by first using five new experimental data. The verification of new model shows that the R-RMSE of corrected model is equal to 0.762%, larger than that of the base model. To further improve the prediction ability of new model, an input and output slope/bias correction is applied on the base model by using all nine new experimental data. If we apply slope/bias correction to all the input variables, totally there are eight parameters to identify. Because of experimental noise, it is not appropriate to identify these parameters with nine experiments. Since the injective velocity has less significant impact on quality, we can ignore it when applying shift/bias correction to input space. Thus, the six parameters of the migrated base model are

**Table 3. Comparison of the Predicted Values from Base Model and Corrected Models**

| data no. | base model (g) | corrected model with OSBC (g) | corrected model with IOSBC (g) | new model (g) | actual value (g) |
|---|---|---|---|---|---|
| 1 | 25.486 | 29.271 | 29.139 | 29.422 | 29.495 |
| 2 | 25.695 | 29.523 | 29.531 | 29.41 | 29.223 |
| 3 | 25.695 | 29.523 | 29.528 | 29.573 | 29.618 |
| 4 | 26.208 | 30.14 | 30.037 | 30.049 | 29.915 |
| 5 | 26.58 | 30.589 | 30.673 | 30.682 | 30.852 |
| 6 | 25.413 | 29.182 | 29.167 | 29.214 | 29.293 |
| 7 | 25.357 | 29.116 | 29.084 | 29.149 | 29.294 |
| 8 | 25.375 | 29.137 | 28.976 | 29.074 | 28.856 |
| 9 | 26.485 | 30.474 | 30.506 | 30.536 | 30.754 |
| 10 | 26.429 | 30.407 | 30.423 | 30.331 | 30.158 |
| 11 | 25.4 | 29.167 | 29.171 | 29.245 | 29.315 |
| 12 | 25.244 | 28.978 | 28.862 | 28.936 | 28.794 |
| 13 | 25.626 | 29.439 | 29.403 | 29.462 | 29.665 |
| 14 | 26.152 | 30.073 | 29.955 | 29.998 | 29.788 |
| R-RMSE | 14.94% | 0.762% | 0.699% | 0.533% | |

identified and the verification result of the new model shows an R-RMSE by 0.699%, still larger than that of the base model. To discover which data have significant prediction error on the corrected base model, an outlier detection algorithm is conducted to guide further DOE for the new process. The four outliers and the corresponding deviation index are found. To design further DOE, new effective radii are specified:

$$r_F = 0.6r = [0.45, 0.21, 0.54] \tag{20}$$

With information of previous centers and outliers taken into account together, another five cluster centers are found by analysis of data generated from the corrected base model. Totally, there are 14 DOE data from the new process. By clustering this data set with larger effective radii, we can roughly divide the whole input space into two local regions. The boundary of the two regions occurs around the middle of process variables. This is consistent with the case where the process variables, especially for packing pressure, have a significant change in impact on quality around the boundary. For the local model structure, the two linear models are applied.

The final two difference models are:

$$R_1: \text{if (BT is } MF_{11}\text{) and (PP is } MF_{21}\text{) and (IV is } MF_{31}\text{), then difference model}_1 = a_{11}\text{BT} + a_{12}\text{PP} + a_{13}\text{IV} + b_1$$

$$R_2: \text{if (BT is } MF_{12}\text{) and (PP is } MF_{22}\text{) and (IV is } MF_{32}\text{), then difference model}_2 = a_{21}\text{BT} + a_{22}\text{PP} + a_{23}\text{IV} + b_2$$

The membership functions are represented by a Gaussian function in eq 17 with found centers ($c_1$= [0.32 0.38 0.26]; $c_2$= [0.68 0.75 0.82]).

The weighted recursive least-squares estimation is conducted to identify the unknown parameters of local difference models. Finally, the new model is obtained by piecing together the two local difference models and the corrected base model. Table 3 summarizes the prediction results of the original base model, corrected base model, and new model built using the model migration techniques discussed in Section II and the actual value. Column two shows the results predicted directly from the base model without any correction. The error is significantly high, suggesting the insufficiency of the base model prediction ability. Columns three and four list the results with the slop/bias correction of the base model. Although the performances with SBC corrected models are significantly better than that without SBC correction, the validation errors are still large. This indicates that the new process is not entirely shifted and rescaled to the old process. The best model is obtained when piecing together the two local difference models and corrected base model (see column five). The model verification result indicates the corresponding MSE is 0.533%, the lowest among all.

## 4. Conclusions

In this paper, we present an economical way of developing a new process model by migrating an existing base model to a new process, using a limited number of experiments. A systematic approach, including process information extracted from a base model, initial design of experiment, outlier detection, further design of experiment, and model migration strategy is proposed. Rather than with a simple case, a comparable complex case is used to demonstrate the proposed method. The new model through the corrected base model plus a global difference model is obtained with fewer data than those required for development of the base model. This shows the proposed method is effective.

## Acknowledgment

This work is supported in part by Hong Kong Research Council, project number 613107, and China National 973 project, 2009CB320603.

## Literature Cited

(1) Meng, J. E.; Liao, J.; Lin, J. Y. Fuzzy Neural Networks-Based Quality Prediction system for Sintering Process. *IEEE Trans. Fuzzy Syst.* **2000**, *8* (3), 314–324.

(2) Lotti, C.; Ueki, M. M.; Bretas, R. E. S. Prediction of the Shrinkage of Injection Molded iPP Plaques Using Artificial Nerual Networks. *J. Injection Molding Technol.* **2002**, *6* (3), 157–176.

(3) Li, E.; Jia, L.; Yu, J. S. A genetic neural fuzzy system and its application in quality prediction in the injection process. *Chem. Eng. Commun.* **2004**, *191* (3), 335–355.

(4) Lee, D. E.; Song, J. H.; Song, S. O.; Yoon, E. S. Weighted Support Vector Machine for Quality Estimation in the Polymerization Process. *Ind. Eng. Chem. Res.* **2005**, *44* (7), 2101–2105.

(5) Lu, N. Y.; Gao, F. Stage-Based Process Analysis and Quality Prediction for Batch Processes. *Ind. Eng. Chem. Res.* **2005**, *44* (10), 3547–3555.

(6) Michael, L.; Thompson, M. A. K. Modeling chemical processes using prior knowledge and neural networks. *AIChE J.* **1994**, *40* (8), 1328–1340.

(7) Van Lith, P. F.; Betlem, B. H. L.; Roffel, B. Combining prior knowledge with data driven modeling of a batch distillation column including start-up. *Comput. Chem. Eng.* **2003**, *27* (7), 1021–1030.

(8) Lu, J.; Gao, F. Process Modeling Based on Process Similarity. *Ind. Eng. Chem. Res.* **2008**, *47* (6), 1967–1974.

(9) Wu, C. F. *Experiments Planning, Analysis, and Parameter Design Optimization*; Wiley: New York, 2000.

(10) Montgomery, D. C. *Design and Analysis of Experiments*, 5th ed.; Wiley; New York, 2001.

(11) Chiu, S. L. Fuzzy model identification based on cluster estimation. *J. Intell. Fuzzy Syst.* **1994**, *2* (3), 267–278.

(12) Meloun, M.; Militky, J. Detection of single influential points in OLS regression model building. *Anal. Chim. Acta* **2001**, *439* (2), 169–191.

(13) Angelov, P. P.; Filev, D. P. An approach to online identification of Takagi-Sugeno fuzzy models. *IEEE Trans. Syst. Man Cybern., Part B* **2004**, *34* (1), 484–498.

(14) Alexandridis, A.; Sarimveis, H.; Bafas, G. A new algorithm for online structure and parameter adaptation of RBF net works. *Neural Networks.* **2003**, *16* (7), 1003–1017.

(15) Takagi, T.; Sugeno, M. Fuzzy identification of systems and its application to modeling and control. *IEEE Trans. Syst. Man Cybern., Part B* **1985**, *15* (1), 116–132.

*Received for review* September 3, 2008
*Revised manuscript received* October 20, 2008
*Accepted* October 22, 2008

IE8013296