A Knowledge-Based Tuning Method for Injection Molding Machines

Dongzhe Yang
Graduate Research Assistant

Kourosh Danai
Professor

David Kazmer
Associate Professor

Department of Mechanical and Industrial Engineering, University of Massachusetts, Amherst, MA 01003

Complexity of manufacturing processes has hindered methodical specification of machine setpoints for improving productivity. Traditionally in injection molding, the machine setpoints are assigned either by trial and error, based on heuristic knowledge of an experienced operator, or according to an empirical model between the inputs and part quality attributes, which is obtained from statistical design of experiments (DOE). In this paper, a Knowledge-Based Tuning (KBT) Method is presented which takes advantage of the a priori knowledge of the process, in the form of a qualitative model, to reduce the demand for experimentation. The KBT Method provides an estimate of the process feasible region (process window) as the basis of finding the suitable setpoints, and updates its knowledge-base using the data that become available during tuning. As such, the KBT Method has several advantages over conventional tuning methods: (1) the qualitative model provides a generic form of representation for linear and nonlinear processes alike, therefore, there is no need for selecting the form of the empirical model through trial and error, (2) the use of a priori knowledge eliminates the need for initial trials to construct an empirical model, so an initial feasible region can be identified as the basis of search for the suitable setpoints, and (3) the search within the feasible region leads to a higher fidelity model of this region when the input/output data from consecutive process iterations are used for learning. The KBT Method's utility is demonstrated in production of digital video disks (DVDs). [DOI: 10.1115/1.1382596]

1 Introduction

Enhancing process efficiency is a continuing objective in manufacturing research. Among various manufacturing processes, injection molding is widely used for producing highly complex and precise plastic parts with tight tolerances and surface finish. This higher precision is often achieved by more accurate control of the process [1,2], which entails, among other things, more accurate specification of machine setpoints such as melt and mold temperature, injection pressure, and injection speed.

The goal of tuning is to specify the setpoints that will yield acceptable part quality attributes. The approaches used for setpoint specification (tuning) are either on-line or off-line. The traditional method of on-line tuning in the plastics industry has been "trial and error." For this, shots are taken during start-up and part quality attributes are measured after each shot to evaluate the quality of the produced parts. The operator then uses his/her knowledge of the process to select the setpoints in such a way as to improve the quality of the part from shot to shot. This tuning exercise is repeated until the specifications for part quality are satisfied. The main drawback of the traditional tuning approach is its inefficiency, because humans usually use linear relationships to relate machine settings to quality attributes, so they often have difficulty adjusting the inputs over large ranges. Furthermore, they tend to treat the various attributes as independent, thus they ignore the couplings among the attributes. With the presence of process and measurement noise, these limitations often result in time-consuming tuning sessions and considerable waste.

Expert systems have also been used for on-line tuning [3–9]. These systems represent the process knowledge in the form of if-then rules, and can sometimes guide the process to the feasible region without invoking too many experimental trials. However, in their current form expert systems are univariate (focus on one problem at a time) and cannot reliably address quality issues not included in the rule-base.

In off-line tuning, an empirical model is developed from a set of designed experiments [10,11]. This model is then used as the basis of a constrained optimization problem to provide the setpoints for the process. Although not as fast as the on-line approach, the off-line approach provides an estimate of the process window, which is important for yield maximization and mold qualification [12–17]. For example, the setpoints may be specified at the center of the process window to provide robustness to process variations. Off-line tuning, however, is often disadvantaged by its demand for extensive experimentation as well as long-term stochastic variation of the process.

In this paper, an on-line method of tuning is introduced that assesses the feasible region of the process as the basis of search for appropriate process setpoints. The novel feature of this method is its incorporation of the a priori knowledge of the process, in the form of a qualitative model, that is updated through learning using the input-output data acquired during tuning. As such, this method offers several advantages over conventional tuning methods. The qualitative model has a generic form, so there is no need for trials to determine the form of the model. Furthermore, this model provides the framework to incorporate linear/nonlinear and monotonic/nonmonotonic sensitivity information between the setpoints and part quality attributes in its initial definition. The KBT Method is on-line (i.e., it interleaves modeling and search), therefore, it always searches for the appropriate setpoints within the estimated feasible region of the model and provides training data for refining this region after each process iteration. This leads to a more concentrated modeling effort within the feasible region and, therefore, a more efficient search of the input-space.

2 The Tuning Method

The strategy used in the Knowledge-Based Tuning (KBT) Method is described in Fig. 1. In this method, the qualitative model is used as the basis of search for the machine setpoints, and

Contributed by the Manufacturing Engineering Division for publication in the JOURNAL OF MANUFACTURING SCIENCE AND ENGINEERING. Manuscript received Jan. 2000; revised Dec. 2000. Associate Editor: R. Furness.

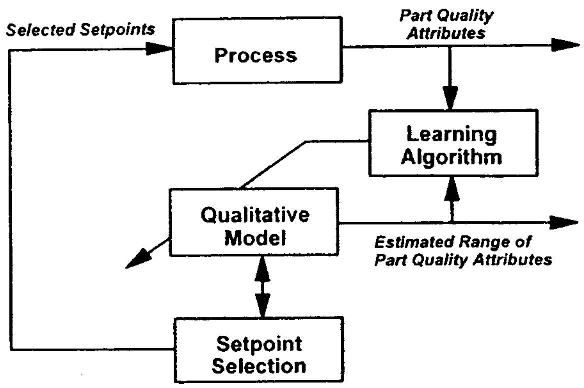

Fig. 1 Schematic of the Knowledge-Based Tuning Method

its accuracy is evaluated by comparing its outputs with the process outputs (i.e., the actual part quality attributes) after each tuning iteration. These evaluation instances are then used to improve the qualitative model, so as to render a better estimation of the feasible region and more suitable machine setpoints.

An earlier version of this tuning strategy was implemented using a quantitative input-output (I-O) model [18]. In that version, input-output data were produced for regression analysis by running the process at several initial arbitrary setpoints. Apart from the inefficiency associated with running trial iterations for modeling, the quantitative model was observed to be influenced strongly by the initial setpoints. The incorporation of a qualitative model in the present version eliminates the need for initial trial cycles for defining the form of the model and its initialization. It also provides robustness to initial setpoints.

2.1 The Qualitative Model. The qualitative knowledge of the process may include causal relationships between the machine settings and various part quality attributes. This knowledge may also contain the level of each machine setting's effect on various part quality attributes. Here, a qualitative model is introduced to represent such knowledge. In this model, the change of each part quality attribute is defined by an interval, as

$$\Delta\overleftrightarrow{y}_j(k)=\overleftrightarrow{C}_{j1}(k)\Delta x_1(k)+\overleftrightarrow{C}_{j2}(k)\Delta x_2(k)+\cdots+\overleftrightarrow{C}_{jn}(k)\Delta x_n(k),\quad j=1,\ldots,m \tag{1}$$

where each parameter denoted with a left-right arrow "↔" is defined as an interval:

$$\overleftrightarrow{C}_{ji}(k)=[C_{Lji}(k),C_{Uji}(k)],\quad i=1,\ldots,n$$

In the above model, $C_{Lji}(k)$ and $C_{Uji}(k)$ represent the current lower and upper bounds of the sensitivity function between each input $\Delta x_i(k)$ and output $\Delta\overleftrightarrow{y}_j(k)$, the interval $\Delta\overleftrightarrow{y}_j(k)$ denotes the estimated range of change of the jth part quality attribute resulted from the change to the current setpoints $[\Delta x_1(k),\ldots,\Delta x_n(k)]$, and k denotes the current process iteration number. The coefficient intervals defined above provide a convenient mechanism for incorporating a variety of qualitative relationships. For example, a monotonically positive relationship can be represented by the coefficient interval $[0,+A]$, or a monotonically negative relationship by the coefficient interval $[-A,0]$, where A is an arbitrary positive number. A coefficient interval $[-A,A]$ will indicate an unknown causal relationship, and the coefficient interval [0, 0] would indicate a null effect. Furthermore, if the relative effect of two machine settings is known, then their upper limits can be defined relative to each other. For example, if the effect of one machine setting on a part quality attribute is known to be at least twice the effect of another machine setting, then the coefficient interval of this machine setting may be set as $[0,+2A]$, defining its upper limit as twice the upper limit of another machine setting having a coefficient interval $[0,+A]$. While the above coefficient intervals are particularly suited to representing qualitative knowledge, they do not preclude exact knowledge. For example, if the effect of the pack pressure on the part dimension is known to be 2 (μm/Mpa), the coefficient interval may be set as [2, 2].

The outputs ranges are obtained according to the following interval analysis computation rules [19]:

$$\overleftrightarrow{C}_{ji}\Delta x_i=[\min\{C_{Lji}\Delta x_i C_{Uji}\Delta x_i\},\max\{C_{Lji}\Delta x_i,C_{Uji}\Delta x_i\}] \tag{2}$$

$$\overleftrightarrow{C}_{ji}\Delta x_i+R=[\min\{C_{Lji}\Delta x_i,C_{Uji}\Delta x_i\}+R,\max\{C_{Lji}\Delta x_i,C_{Uji}\Delta x_i\}+R] \tag{3}$$

where R denotes a constant, and

$$\overleftrightarrow{C}_{j1}\Delta x_1+\overleftrightarrow{C}_{j2}\Delta x_2=[\min\{C_{Lj1}\Delta x_1,C_{Uj1}\Delta x_1\}+\min\{C_{Lj2}\Delta x_2,C_{Uj2}\Delta x_2\},\times\max\{C_{Lj1}\Delta x_1,C_{Uj1}\Delta x_1\}+\max\{C_{Lj2}\Delta x_2,C_{Uj2}\Delta x_2\}] \tag{4}$$

Once the range of change of each output is computed, the estimated range of the output $\overleftrightarrow{y}_j(k)$ at any potential set of inputs $[\bar{x}_1,\bar{x}_2,\ldots,\bar{x}_n]$ can be computed relative to a reference input set $[x_1(k),x_2(k),\ldots,x_n(k)]$ and its associated output $y_j(k)$ as

$$\begin{aligned}\overleftrightarrow{y}_j(k)&=y_j(k)+\Delta\overleftrightarrow{y}_j(k)\\&=y_j(k)+\overleftrightarrow{C}_{j1}\Delta x_1(k)+\cdots+\overleftrightarrow{C}_{jn}\Delta x_n(k)\\&=[y_j(k)+\min\{C_{Lj1}\Delta x_1(k),C_{Uj1}\Delta x_1(k)\}+\cdots\\&\quad+\min\{C_{Ljn}\Delta x_n(k),C_{Ujn}\Delta x_n(k)\},y_j(k)\\&\quad+\max\{C_{Lj1}\Delta x_1(k),C_{Uj1}\Delta x_1(k)\}+\cdots\\&\quad+\max\{C_{Ljn}\Delta x_n(k),C_{Ujn}\Delta x_n(k)\}]\end{aligned} \tag{5}$$

where

$$\Delta x_i(k)=\bar{x}_i-x_i(k),\quad i=1,\ldots,n \tag{6}$$

represents the difference between the potential input and the reference input. A reference input denotes an input for which the actual value of the output is available to provide the reference for estimating the output range at other potential inputs. In application to injection molding, a reference input is one that has been applied to the process, thus, its suitability has been explored. Hereafter, "reference" and "explored" inputs will be used interchangeably for inputs that have already been applied to the process.

In order to provide insight about the representation capability of the qualitative model, the fit provided by the interval model for a nonlinear single-input single-output relationship is illustrated in Fig. 2, with the output range estimated relative to one reference input. Although the output range is quite large at points far from the reference input in Fig. 2, it still encompasses the actual output, mainly due to the model's compliance with the actual input-output relationship and its conservatively defined coefficient interval. According to Eq. (1), the estimated range of the output becomes larger, and therefore less accurate, as the potential input is selected farther from the reference input to yield a larger Δx_i. This potential drawback of the interval model is eliminated once several reference inputs become available for estimating the output range.

When several reference inputs are available, the estimated output $\overleftrightarrow{y}_j(k)$ at a potential input $\bar{x}_i$ may be computed relative to any set of previously explored inputs, yielding different estimates of $\overleftrightarrow{y}_j(k)$ (due to different values of $\Delta x_i(k)$). In order to cope with the multiplicity of estimates, $\overleftrightarrow{y}_j(k)$ is defined here as the common range among all of the $\overleftrightarrow{y}_j(k)$ estimates. The estimation of $\overleftrightarrow{y}_j(k)$ using this commonality rule is illustrated in Fig. 3 when $\overleftrightarrow{y}_j(k)$ is estimated relative to five reference inputs. It is clear from Fig. 3

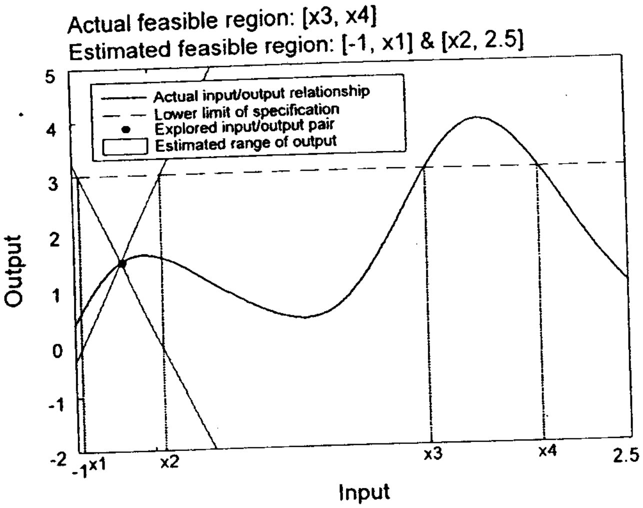

Fig. 2 Estimated range of output by the interval model based on one reference input

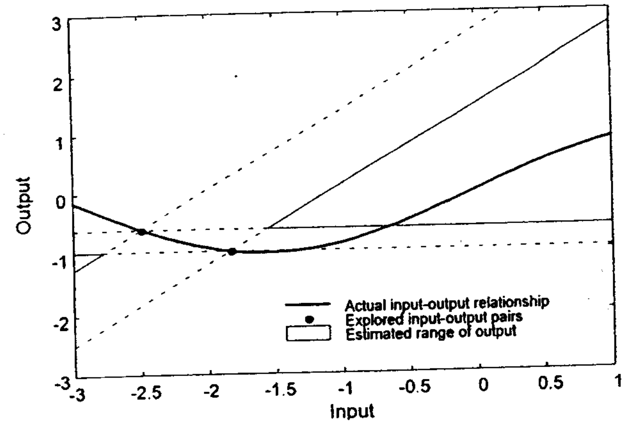

Fig. 4 Estimated range of output by a monotonic interval model for a nonmonotonic input-output relationship

that using this estimation approach for $\tilde{y}_j(k)$ enables representation of the system nonlinearities in a piece-wise fashion. This ability of the qualitative model in representing linear and nonlinear relationships alike makes it ideally suited to modeling ill-defined processes. Of course, this representation capability of the interval model is only successful when the model coefficients comply with the overall monotonicity between individual inputs and outputs. A case where such compliance does not exist is illustrated in Fig. 4, where the lack of commonality between the estimated ranges of the output result in a part of the input-output relationship not to be represented by the interval model. Any lack of compliance between the interval model and the input-output relationships can be corrected by adapting the coefficient intervals through learning.

2.2 Feasible Region Estimation. In order to estimate the feasible region, the individually estimated $\tilde{y}_j(k)$ are compared with their corresponding part quality specifications, so as to speculate whether the corresponding machine setpoints belong to the feasible region or not. In its present form, even when the interval $\tilde{y}_j(k)$ only partly overlaps the part quality specification, the corresponding machine setpoints are included in the estimated

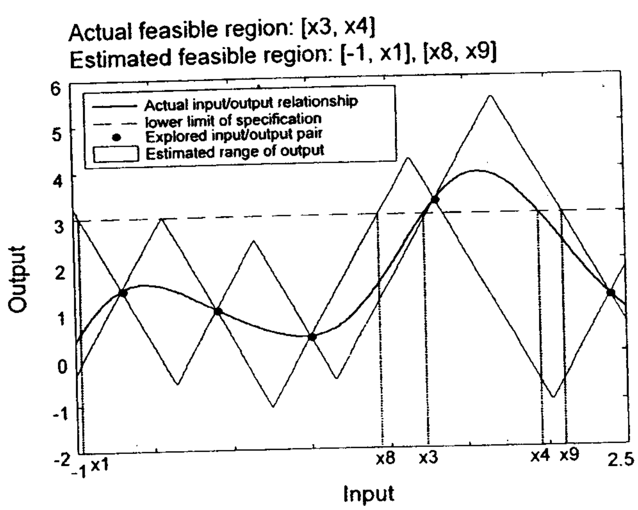

Fig. 3 Estimated range of output by the interval model based on five reference inputs

feasible region. As was discussed earlier, the estimated feasible region is the basis for setpoint selection. As such, it is crucial that the method at least partially estimates the actual feasible region. A partial estimation of the feasible region will ensure consideration of some of the appropriate setpoints during the search, whereas a complete misrepresentation of it will lead to failure in tuning. A complete representation of the feasible region will ensure optimization of the process.

The estimation of the feasible region for a one-sided constraint of $y \geqslant 3$ is illustrated in Figs. 2 and 3. As shown in these figures, all the input ranges that correspond to $\tilde{y} \geqslant 3$ are included within the estimated feasible region, i.e., the range $[-1, x_1]$ and $[x_2, 2.5]$ in Fig. 2, and the range $[-1, x_1]$ and $[x_8, x_9]$ in Fig. 3. The actual feasible region, which includes the inputs associated with $y \geqslant 3$, comprises the range $[x_3, x_4]$. The estimated feasible region, although larger than the actual one, consistently encompasses the actual feasible region, and that the estimated feasible region converges towards the actual region as more reference inputs become available. The consistent estimation of the feasible region in this case is due to the conservative definition of the interval model to represent the actual input-output relationship at all inputs. While conservatism improves the method's likelihood of finding the optimal settings, it leads to an overestimation of the feasible region and slower convergence. We rely on learning to tighten the interval model to accelerate convergence. The convergence of the estimated feasible region towards the actual feasible region with more reference inputs, also illustrated in these figures, is an important factor in the efficiency of the KBT Method. A smaller feasible region means a smaller number of potential input sets to consider, so a rapidly shrinking region becomes synonymous with a faster tuning session. The above procedure of estimating the feasible region based on a single output is then extended to multiple outputs by forming the conjunction of the estimated feasible regions from each output.

2.3 Learning. Although the a priori knowledge may provide a good initial estimate of the interval model, it may not be able to carry the search process to the end unless it is validated and refined through learning. In the proposed method, the coefficient intervals of the interval model are updated according to the input-output data acquired during tuning by considering new values for each of the upper and lower limits of individual coefficient intervals. First, the output ranges at each of the previously explored setpoints are estimated according to the current interval model so that the adapted model will accurately represent all of the previously explored outputs. The adaptation of the coefficient intervals may be performed in two steps: enlargement and shrinkage. If the estimated output ranges do not include the actual outputs, the upper and lower limits of the first coefficient interval will be enlarged by a small step to yield an enlarged coefficient interval, as

$$[C_{Lji}(k)-\beta, C_{Uji}(k)+\beta] \quad (7)$$

where the adaptation step β is defined proportional to the current interval length, as

$$\beta=\gamma[C_{Uji}(k)-C_{Lji}(k)] \quad (8)$$

The parameter γ is usually selected within the range [1/50, 1/20]. After this enlargement, the output ranges are re-estimated using the enlarged coefficient. If the newly estimated output ranges do not include the actual outputs, the next coefficient interval will be enlarged, and its success in improving the estimation capability of the model will be evaluated. Enlargement of the coefficient intervals will be repeated one at a time until the estimated output ranges include all of the actual outputs.

Although the newly updated interval model will enclose the actual outputs after the enlargement phase, it may have coefficient intervals that are overly broad. In order to rectify this situation, the coefficient intervals will be shrunk individually by selecting two new candidates, respectively, for their upper and lower limits in proportion to the current interval length. The two new candidate coefficient intervals will have the form

$$[C_{Lji}(k)+\beta, C_{Uji}(k)] \text{ and } [C_{Lji}(k), C_{Uji}(k)-\beta] \quad (9)$$

and their suitability will be tested by how well they include the actual outputs within the newly estimated ranges. The goodness of fit of the outputs for each candidate interval is quantified by the evaluation function

$$E=\sum_{j=1}^{m}\sum_{q=1}^{k}\{[y_j(q)-y_{jl}(q)]^2+[y_{ju}(q)-y_j(q)]^2\} \quad (10)$$

where m represents the number of outputs, k represents the current process iteration, $y_{jl}(q)$ and $y_{ju}(q)$ represent, respectively, the lower and upper limits of the estimated output ranges, and $y_j(q)$ denotes the actual output. A smaller evaluation function E indicates a more centered set of outputs within the estimated ranges, which is more desirable, so the candidate interval with a smaller value of E is selected. This procedure of shrinking coefficients will be repeated for every coefficient interval in the model. Although the univariate shrinkage of the coefficient intervals will improve the centrality of the model, it will not necessarily optimize it. The shrinkage procedure is repeated in a cyclic fashion until either the estimated output ranges begin to exclude the actual outputs or the intervals violate the minimum length constraint defined as

$$\min L=\{C_{Uji}(0)-C_{Lji}(0)\}(1-\alpha)^k \quad (11)$$

to prevent drastic shrinkage of the coefficient intervals. In the above limit, the parameter $\alpha\in[0,1]$ determines how fast the coefficient interval will be shrunk, and the parameter k denotes the current process iteration. The coefficient interval cannot be shrunk when $\alpha=0$ and can be shrunk without limits when $\alpha=1$. With α equal to 0.1, a coefficient interval can be shrunk to 35 percent of its original length at the 10th iteration and to 12 percent at the 20th iteration. For conservatively defined initial coefficient intervals, the parameter α can be set to a large value (e.g., between 0.15 and 0.3) in order to have a fast shrinkage rate. Conversely, for a system with several inputs, the parameter α may be set to a small value (e.g., between 0.05 and 0.15), since a slow shrinkage rate is preferred due to the larger number of data needed for training. A small value of α is also preferred for a highly nonlinear process which needs extensive training. In general, the value of α only affects learning at the beginning of the tuning process, because once enough data becomes available learning will be driven more by the data than by the value of α. The performance of the learning algorithm is demonstrated in correcting the inconsistency

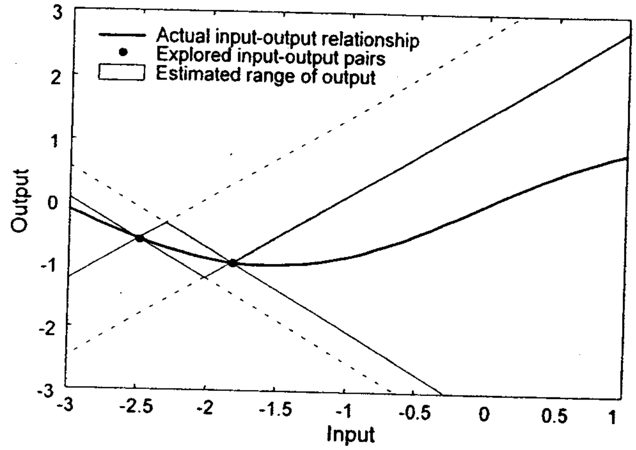

Fig. 5 Estimated range of output by the updated interval model for a nonmonotonic input-output relationship

of the interval model in Fig. 4 based on the two reference inputs. The estimated outputs by the updated interval model, shown in Fig. 5, indicate the effectiveness of the learning algorithm in correcting initially ill-defined models.

2.4 Characterization of Part Quality. The quality of part attributes is the performance measure for the KBT Method. This measure is quantified in terms of the relative distance of the part attributes from their upper and lower specifications. Accordingly, a perfectly tuned process should produce molded parts with quality attributes at the center of their corresponding allowable ranges and having very small variances. The metric for part quality can be defined by the standard deviation number (λ_j) for each quality attribute in terms of the mean and standard deviation of the quality attribute, as

$$\lambda_j=\text{Min}\left(\frac{Y_{ju}(0)-\mu_j}{\sigma_j},\frac{\mu_j-Y_{jl}(0)}{\sigma_j}\right), \quad (12)$$

where $Y_{ju}(0)$ and $Y_{jl}(0)$ denote, respectively, the upper and lower specifications of the jth quality attribute, and μ_j and σ_j represent, respectively, the mean and standard deviation of the part attribute. According to this definition, a larger standard deviation number implies a more centered quality attribute within its range and/or a more consistent quality attribute. In this study, the smallest standard deviation number (SSDN) λ_s of the quality attributes, which represents the distance of the worst quality attribute from its process defect boundary, is considered the evaluation metric for the quality of produced parts. According to Gaussian distribution, a smallest standard deviation number equal to 3 corresponds to a nominal yield of 99.73 percent.

2.5 Selection of the Machine Setpoints. The machine setpoints selected within the estimated feasible region may actually be infeasible due to the inaccuracy of the estimated region. This will lead to unacceptable parts. In order to enhance the quality of the produced parts, the reliability of the selection region may be increased beyond the estimated feasible region by using a more aggressive set of part quality specifications to define the selection region as

$$Y_{ju}(k)=Y_{ju}(0)-\sigma_j(l)*\lambda_s(l) \quad (13)$$

$$Y_{jl}(k)=Y_{jl}(0)+\sigma_j(l)*\lambda_s(l) \quad (14)$$

where

$$l=\arg\max_q\{\lambda_s(q)\} \quad q=1,\ldots,k \quad (15)$$

In the above formulation, $Y_{ju}(k)$ and $Y_{jl}(k)$ denote the tighter upper and lower specifications of the jth part quality attribute at the current process iteration k, $Y_{ju}(0)$ and $Y_{jl}(0)$ represent the original upper and lower specifications, respectively, $\lambda_s(l)$ represents the smallest standard deviation number of the parts at the lth process iteration, $\sigma_j(l)$ denotes the standard deviation of the jth part quality attribute at the lth iteration, and l represents the process iteration that produced the largest smallest standard deviation number for the part quality attributes so far. These closer specifications can then be used to define the selection region for the machine setpoints using the same approach as feasible region estimation. The analogy behind the above formulation is that given the lth process iteration being the best cycle so far from the point of view of yield, the selection region is defined inside the estimated feasible region so as to encompass the setpoints that may produce parts with a larger λ_s than that at the lth iteration. This selection region will then include the candidate setpoints for the next process iteration.

Apart from producing good parts, the setpoints need to be selected so as to provide sufficient exploration of the input space for feasible region estimation. As such, setpoint selection becomes synonymous with increasing the distance of the selected setpoints from the previous setpoints, while having them close to the center of the selection region. This objective can be pursued by minimizing the following objective function:

$$S=\frac{\sum_{e=1}^{N_e}\text{Distance}(\mathbf{x}_c,\mathbf{x}_e)}{\left(\prod_{s=1}^{N_s}\text{Distance}(\mathbf{x}_c,\mathbf{x}_s)\right)^{1/N_s}}, \tag{16}$$

where $\mathbf{x}_c$ represents a candidate set of setpoints within the selection region, $\mathbf{x}_e$ represents any candidate setpoint set within the selection region, $\mathbf{x}_s$ denotes each of the previously explored setpoints, and N_e and N_s represent the number of the estimated feasible setpoints and the previously explored setpoints, respectively. Note that when the candidate set of setpoints $\mathbf{x}_c$ is close to the previously selected setpoints, $(\Pi_{s=1}^{N_s}\text{Distance}(\mathbf{x}_c,\mathbf{x}_s))^{1/N_s}$ becomes small, and when the candidate set of setpoints $\mathbf{x}_c$ is far from the center of the selection region, the value of $\Sigma_{e=1}^{N_e}\text{Distance}(\mathbf{x}_c,\mathbf{x}_e)$ becomes large. By minimizing S, the candidate setpoints are selected such that a balance between the above two objectives is obtained.

3 Benchmarking

The performance of the KBT Method was first studied in simulation, where the complexity of the test domain could be controlled and the feasible region could be explicitly mapped in the input space for the purpose of evaluating the estimation capability of the KBT Method. The two inputs of the Test Model represent pack pressure (MPa) and shot size (mm), and its three outputs denote the two qualitative attributes of short shot (ratio) and flash (ratio), and one quantitative dimensional attribute (mm). The Test Model, that is defined to represent empirical relationships between the corresponding inputs and outputs, has the form

$$y_1=\frac{0.2}{0.2+x_2-50}-0.0196 \tag{17}$$

$$y_2=\frac{1.0}{1+e^{0.5(110-x_1)}} \tag{18}$$

$$y_3=117\left(1-\frac{0.01(120-x_1)}{40}+0.0005x_2\right) \tag{19}$$

For simulation purposes, the input domain for the Test Model was set as $X_u=[120,60]$ and $X_l=[80,50]$, and the upper and lower quality specifications were set as $Y_u=[0.015,0.015,119.75]$ and $Y_l=[0.0,0.0,119.25]$, respectively. In order to take process uncertainty into consideration, the pack pressure and shot size were perturbed with normally distributed noise levels with standard deviations equal to 1.0 percent and 0.4 percent of their respective absolute values. The part quality attributes of the Test Model over the range of its inputs are shown in Fig. 6, along with its feasible region.

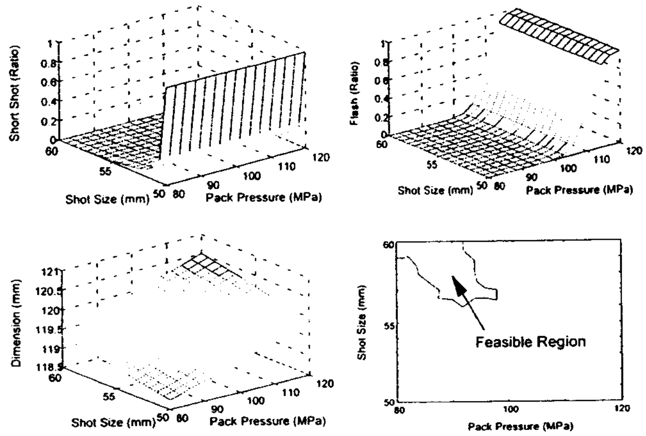

Fig. 6 Part quality attributes and the feasible region of the Test Model

The estimation capability of the feasible region by the KBT Method was tested first. For this, a two-input three-output qualitative model was defined as

$$\Delta\vec{y}_1(k)=\vec{C}_{11}\Delta x_1(k)+\vec{C}_{12}\Delta x_2(k) \tag{20}$$

$$\Delta\vec{y}_2(k)=\vec{C}_{21}\Delta x_1(k)+\vec{C}_{22}\Delta x_2(k) \tag{21}$$

$$\Delta\vec{y}_3(k)=\vec{C}_{31}\Delta x_1(k)+\vec{C}_{32}\Delta x_2(k) \tag{22}$$

where each coefficient interval was defined according to the overall sensitivity between the corresponding input and output, as

$$\vec{C}_{11}=[0,0],\quad \vec{C}_{12}=[-4,0] \tag{23}$$

$$\vec{C}_{21}=[0,4],\quad \vec{C}_{22}=[0,0] \tag{24}$$

$$\vec{C}_{31}=[0,4],\quad \vec{C}_{32}=[0,4] \tag{25}$$

Tuning was initiated with a random set of initial setpoints, at which ten outputs were generated to simulate the stochastic behavior of the process at each setpoint set. Learning began with the second process simulation, with the parameter α in Eq. (11) set to 1/7. Based on the outputs from each process simulation, the feasible region was estimated and the setpoints for the next process

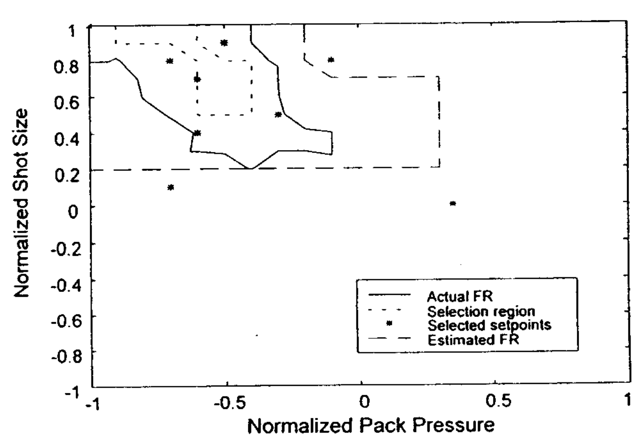

Fig. 7 The estimated and selection regions by the KBT Method for the Test Model after 9 iterations

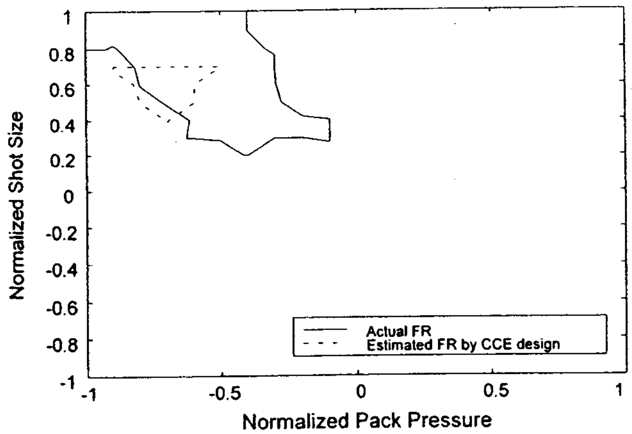

Fig. 8 The estimated feasible region by CCE design for the Test Model

iteration were selected within the selection region by minimizing the objective function S in Eq. (16). The above procedure was repeated through nine iterations. The estimated feasible region and selection region are shown in Fig. 7 relative to the actual feasible region. The results indicate that the estimated feasible region, although broader than the actual feasible region, encompasses the proper region of the input space and completely includes the actual feasible region. This is encouraging, since an important feature of the KBT Method is to provide a reliable estimate of the process window during tuning. The results in Fig. 7 also indicate that the selection region, as defined by a stringent set of part quality specifications, is completely within the feasible region. This behavior ensures consideration of the appropriate inputs by the KBT Method for the next process iteration. Persistent positioning of the selection region within the actual feasible region is key to successful optimization of the process by the KBT Method.

The estimated feasible region was next compared with one obtained from a DOE method. For this, the method of Central Composite Experimental (CCE) design was used, which is a two-level fractional design augmented with a small number of data to permit estimation of a second-order regression model. Nine sets of setpoints were evaluated according to CCE design. The means and standard deviations of these part quality attributes were then used to train second-order regression models for estimating the smallest standard deviation number (SSDN) of the parts' quality attributes at various setpoints. The set of setpoints within the input-space that generated SSDNs greater than or equal to 3 were considered inside the feasible region. The estimated feasible region by CCE design is shown in Fig. 8. The results indicate that although there is a significant overlap between the estimated feasible region by the CCE design and the actual feasible region, the estimate only includes a portion of the actual feasible region. This limits the capability of the CCE design to find an optimal setpoint for the process.

4 Experimental Validation

The practical utility of the Knowledge-Based Tuning Method was tested in production of digital video disks (DVD) molded of a commercial optical grade polycarbonate. The DVD geometry consists of a center-gated disk of 60 mm radius and 0.6 mm thickness. Because the DVD is used in a high-density optical storage application, it is necessary to minimize residual stress, birefringence, flatness, and track-to-track dimensional deviation (groove replication). The KBT Method was tested in selecting five of the machine settings on a Sumitomo SD30 molding machine. The machine settings included the melt barrel temperature, the mold coolant temperature, the first stage clamp tonnage, the second stage clamp tonnage, and the first stage clamp tonnage delay. The

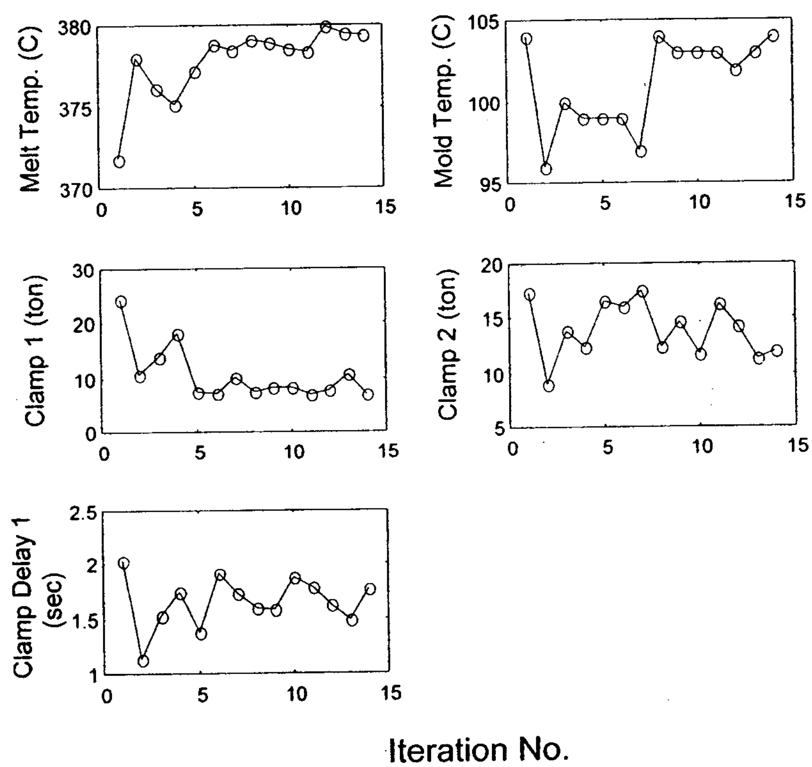

Fig. 9 The machine setpoints selected by the KBIM Method at each process iteration

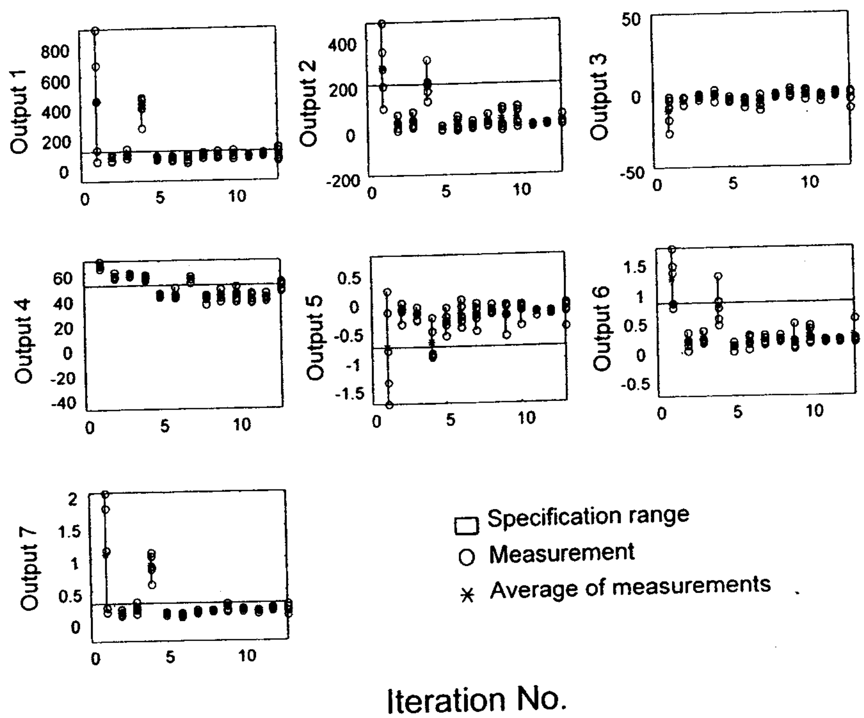

Fig. 10 Measurements of the part quality attributes produced by the selected machine setpoints. Output 1: OD Deviation (micron), Output 2: Vertical Deviation (micron). Output 3: Min. Birefringence (nm), Output 4: Max. Birefringence (nm), Output 5: Min. Radial Deviation (degree), Output 6: Max. Radial Deviation (degree), Output 7: Max. Tangential Deviation (degree).

upper and lower bounds for the machine settings were set to the global processing limits that are conventionally used for optical molding, as $X_u=[380°\text{C},104°\text{C},25\text{ ton},18\text{ ton},2.2\text{ sec}]$ and $X_l=[370°\text{C},95°\text{C},7\text{ ton},8\text{ ton},1\text{ sec}]$.

Among the part quality attributes commonly considered for DVDs, seven critical attributes were considered: *OD deviation, vertical deviation, minimum and maximum birefringence* levels across the molded DVDs, *minimum and maximum radial deviation, and maximum tangential deviation* from a flat surface. The upper and lower specifications for the seven part quality attributes were selected as $Y_u=[100$ micron, 200 micron, 50 nm, 50 nm, 0.8 deg, 0.8 deg, 0.3 deg$]$ and $Y_l=[-100$ micron, -200 micron, -50 nm, -50 nm, -0.8 deg, -0.8 deg, -0.3 deg$]$. The experimental procedure at each new machine setting consisted of allowing the new melt and mold temperatures to reach their steady states, discarding the first twenty molded disks, and then taking part quality measurements from the next five molded disks. Disks were allowed to cool for approximately 15 to 25 minutes before their part quality attributes were measured by an electronic scale, an electronic micrometer, and a specialized optical media measurement instrument (Dr. Schenk Prometheus).

The selection region was defined to include those inputs that generate smallest standard deviation numbers larger than 3. This would mean that the estimated output range by the interval model, $\tilde{y}_j$, would be compared against a more stringent specification range defined as $[Y_{jl}+3\hat{\sigma}_j, Y_{ju}-3\hat{\sigma}_j]$. These specifications were defined based on standard deviation estimates from DOE experiments of another material, as: $\hat{\sigma}=[9.6162, 3.4718, 5.0608, 4.1174, 0.0235, 0.0238, 0.0206]$ [20].

Seven interval models were defined to estimate the range of the seven part quality attributes. For a priori knowledge, the range of sensitivity coefficients from DOE models of a different commercial optical grade polycarbonate were used. Tuning began with a set of machine setpoints as $X=[371.8°\text{C}, 104°\text{C}, 24.4\text{ ton}, 17.3\text{ ton}, 2.03\text{ sec}]$, and learning was activated after the second iteration, when values for Δx became available. For learning, the parameter α was set to 0.1. Altogether 13 tuning iterations were performed. The machine setpoints used at each iteration are shown in Fig. 9, and the part quality attributes at these machine setpoints are shown in Fig. 10. The results indicate that six of the seven part quality attributes initially violated their corresponding specifications, and that at the fifth iteration all of the part quality attributes were within their specifications. Given the large number of inputs and outputs considered in these experiments, the speed of tuning provided by the KBT Method is remarkable. This same process to be tuned by the CCE design, for example, would require 45 iterations.

Although the parts produced after the 5th iteration were within their specifications, they were not necessarily robust. The quality of the produced parts at each iteration are indicated by the small-

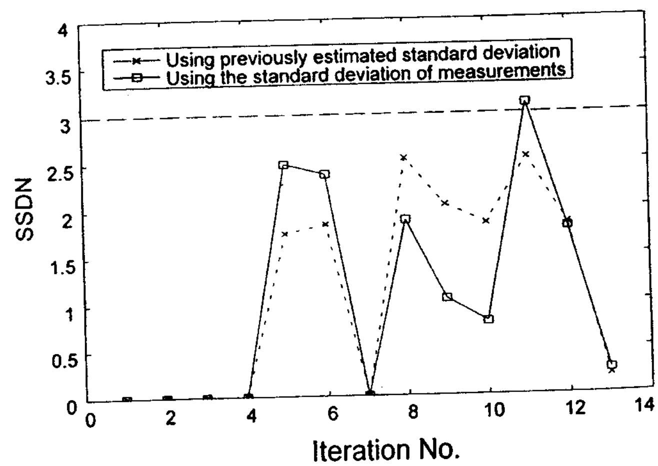

Fig. 11 Smallest standard deviation numbers (SSDN) calculated using two sets of standard deviation values

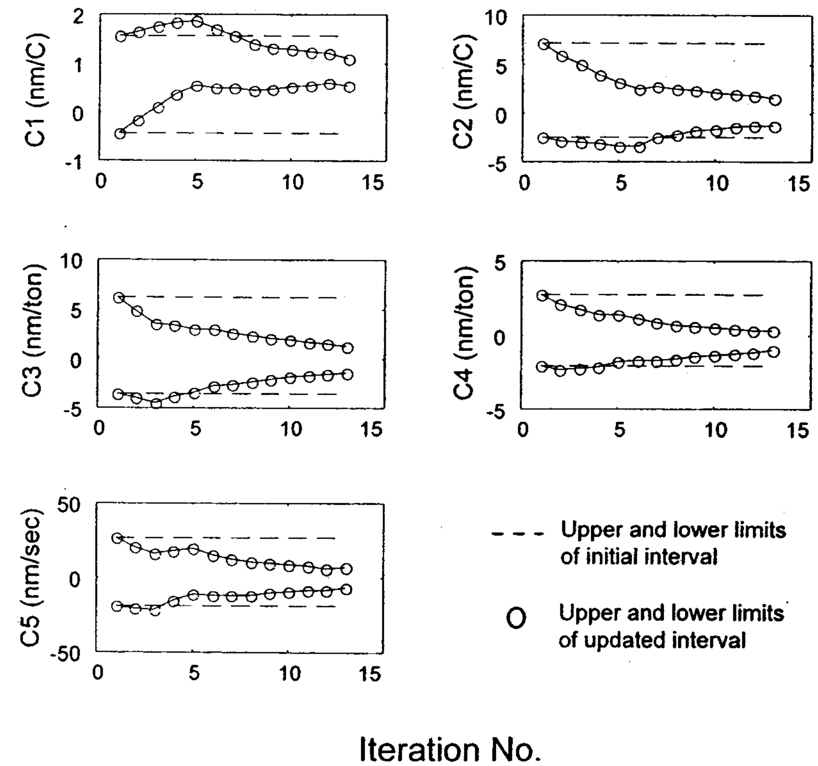

Fig. 12 The initial and updated coefficient intervals of the minimum birefringence during the experiments

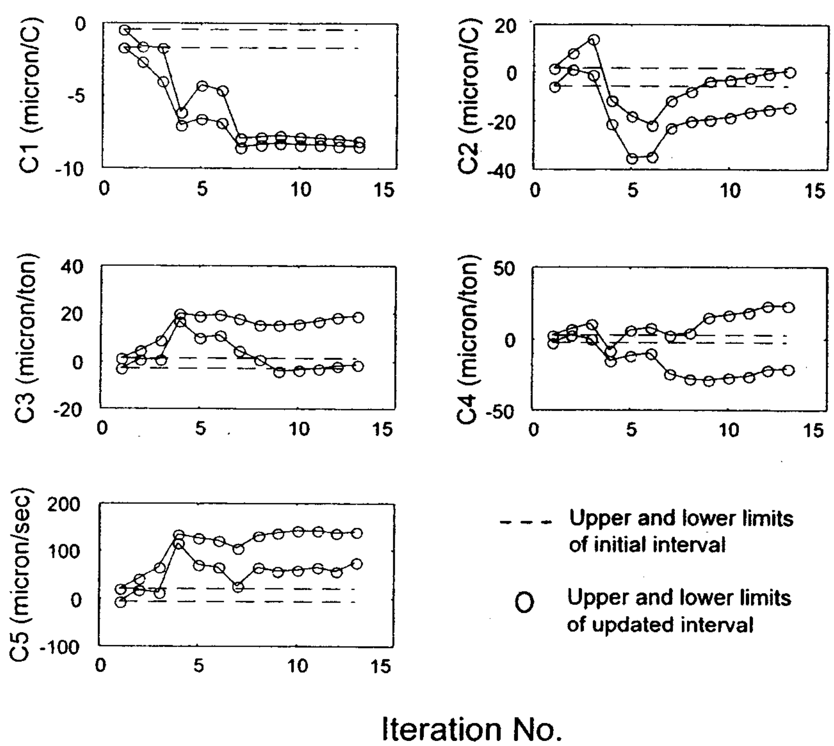

Fig. 13 The initial and updated coefficient intervals of the vertical deviation during the experiments

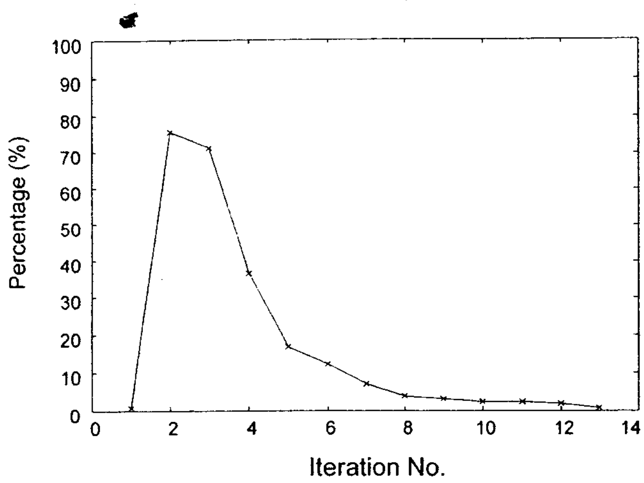

Fig. 14 Percentage of the selection region in the search space

est standard deviation number (λ_s values) of their attributes in Fig. 11. These λ_s values were obtained using two sets of standard deviation values. One set of standard deviations was estimated from DOE models of a different optical grade polycarbonate [20], and the second set was obtained in cycle from the part quality attribute measurements at each iteration. The results in Fig. 11 indicate that although the smallest standard deviation numbers of the part quality attributes increased progressively with each iteration until iteration 11, the smallest standard deviation number never exceeded the target value of 3. According to the λ_s values from the measurements, the selected machine setpoints at the 11th iteration barely satisfied the robustness criterion, and the setpoints at iterations 12 and 13 produced parts with inferior quality than iteration 11. The relative divergence of the setpoints from their progressively improving trend is perhaps due to a divergence of the estimated selection region from the actual feasible region. In practice, the occurrence of such divergence will be detected from a deterioration in the quality of produced parts. One remedy is to reset the qualitative model to the state that has produced the best parts (iteration 11 in this case) and continue with training without implementing the recommendations of the method for setpoint adjustment, until it becomes necessary to re-tune the process due to a drift in the quality of produced parts.

The role of learning in these experiments is illustrated through the values of coefficient intervals of the *minimum birefringence* and the *vertical deviation* shown in Figs. 12 and 13, respectively. The results in Fig. 12 indicate that the updated coefficient intervals are refined within their initial bounds, whereas the updated coefficient interval values in Fig. 13 indicate that the initial bounds were not accurate and that the intervals needed to be moved beyond the initial bounds. These updated coefficient intervals also indicate monotonic relationships defined by C_1 (from the *melt barrel temperature*) and C_5 (from the *first stage clamp tonnage delay*), as evidenced by the negative and positive coefficient intervals C_1 and C_5, respectively.

Another measure of convergence of the KBT Method is the size of the selection region relative to the initial search space. The ratio of the selection region at each tuning iteration relative to the original search space is shown in Fig. 14. The results indicate that the estimated selection region was very small at the first iteration because the initial interval model was not accurate. After the interval model was updated at the second iteration, it produced a fairly large selection region approximately equal to 75 percent of the search space. The machine setpoints selected within such a large selection region could easily fall out of the actual feasible region, as evidenced by the unacceptable parts produced at the second, third, and fourth iterations. However, when more experimental data were available for updating the interval model, the selection region began to converge to the actual feasible region. For example, at the fifth iteration, the selection region was smaller than 20 percent of the search space, and coincided well with the actual feasible region, which included the machine setpoints selected at the fifth and sixth iterations. The best set of machine setpoints was selected at the 11th iteration where the size of the selection region was only 2.4 percent of the search space.

The results indicate that the KBT Method succeeded in finding setpoints that produced acceptable parts within 5 process iterations, and that the process could not be operated at the defined level of robustness. Two underlying assumptions are made at the outset of any tuning session: (1) that a feasible region exists to be found, and (2) that the machine capability is adequate in controlling the quality of the part (i.e., controllability). The truth, however, is that one or both of these assumptions may be invalid. Although the success or failure of the tuning method provides an indication of validity of these underlying assumptions, it is important that the tuning method be capable of identifying process limitations. Our future development efforts will include the addition of such capabilities to the KBT Method.

5 Conclusion

An on-line method of tuning is introduced for injection molding that incorporates the a priori knowledge of the process in its initial formulation. The salient features of this on-line tuning method is its continual estimation of the process window during tuning, and use of learning to adapt its input-output model after each tuning iteration. The performance of this KBT Method in estimating the process window is demonstrated in simulation and evaluated experimentally in production of digital video disk (DVD) substrates.

6 Acknowledgments

The authors would like to thank GE plastics for providing access to the Optical Media Development Center. Portions of this work were supported by the National Science Foundation (Grants No. DDM-9114484 and DMI-9700288).

References

[1] Agrawal, A. R., and Pandelidis, I. O., 1987, "Injection-Molding Process Control-a Review," Polym. Eng. Sci., **27**, pp. 1345–1357.
[2] Amellal, K., Tzoganakis, C., Penlidis, A., and Rempel, G. L., 1994, "Injection Molding of Medical Plastics: a Review," Adv. Polym. Technol., **13**, pp. 315–322.
[3] Farrell, R. E., and Dzeskievicz, L., 1994, "Expert Systems for Injection Molding," *SPE ANTEC Conference Proceedings*.
[4] Kameoka, S., Haramoto, N., and Sakai, T., 1993, "Development of an Expert System for Injection Molding Operations," Adv. Polym. Technol., **12**, pp. 403–418.
[5] Rogers, J. K., 1991, "Intelligent' Molding: Expert Systems Are Coming on Line Now," Mod. Plast., **68**, pp. 56–60.
[6] Jong, W. R., and Hsu, S. S., 1997, "An Integrated Expert System for Injection-Molding Process," *SPE ANTEC Conference Proceedings*.
[7] Jan, T. C., and O'Brien, K. T., 1993, "A User-friendly, Interactive Expert System for the Injection Molding of Engineering Thermoplastics," International Journal of Advanced Manufacturing Technology, **8**, pp. 42–51.
[8] Pandelidis, I. O., and Kao, J. F., 1990, "DETECTOR, a Knowledge-Based System for Injection Molding Diagnostics," Journal of Intelligent Manufacturing, **1**, pp. 49–58.
[9] Shelesh-Nezhad, K., and Siores, E., 1997, "An Intelligent System for Plastic Injection Molding Process Design," J. Mater. Process. Technol., **63**, pp. 458–462.
[10] Schmidt, S., and Launsby, R., 1988, *Understanding Industrial Designed Experiments*, Air Academy Press, Colorado Springs, Colorado.
[11] Neter, J., Kutner, H. M., Nachtsheim, C. J., and Wasserman, W., 1996, *Applied Linear Statistical Models*, IRWIN, 4th ed.
[12] Phadke, M., 1989, *Quality Engineering Using Robust Design*, Prentice Hall, Englewood Cliffs, New Jersey.
[13] Budill, K., 1993, "A Systematic Approach to Tool Qualification for Injection Molding," Master's thesis, Massachusetts Institute of Technology, Boston, MA.
[14] Martin, M. F., Bontumasi, F., and Young, G., 1995, "The Practical Application of Design of Experiments in the Total Quality Injection Molding Process," *SPE ANTEC Conference Proceedings*, Boston, MA.
[15] Michaeli, W., Vaculik, R., and Bluhm, R., 1995, "Reprocessing of Recyclat-on-Line Prediction of Quality Attributes," *SPE ANTEC Conference Proceedings*, Boston, MA.
[16] Michaeli, W., and Vaculik, R., 1995, "Closed Loop Quality Control for In-

jection Molding based on Statistical Process Models," *SPE ANTEC Conference Proceedings*, Boston, MA.
[17] Vaatainen, O., Jarvela, P., Valta, K., and Jarvela, P., 1994, "The Effect of Processing Parameters on the Quality of Injection Moulded Parts by Using the Taguchi Parameter Design Method," Plast. Rubber Process. Applic., **21**, pp. 211–17.
[18] Ivester, R., and Danai, K., 1998, "Tuning and Automatic Regulation of Injection Molding by the Virtual Search Method," ASME J. Manuf. Sci. Eng., **120**, pp. 323–329.
[19] Moore, R. E., 1979, *Methods and Applications of Interval Analysis*, Society for Industrial and Applied Mathematics, Philadelphia.
[20] Hatch, D., 1999, "Transfer Function Development for the Injection Molding of Optical Media," Master's thesis, University Of Massachusetts, Amherst, MA.