Recursive Data-Based Prediction and Control of Batch Product Quality

S. A. Russell, P. Kesavan, and J. H. Lee
Dept. of Chemical Engineering, Auburn University, Auburn, AL 36849

B. A. Ogunnaike
Experimental Station, E1/102, E. I. DuPont de Nemours and Co., Wilmington, DE 19880

In typical batch and semibatch processes, process/feedstock disturbances occur frequently and on-line measurements of product quality variables are not available. As a result, most batch processes have not been able to achieve tight quality control. Empirical, data-driven approaches are very attractive for dealing with this problem because of the difficulties associated with developing accurate process models from first principles. An approach for recursive on-line quality prediction was developed around data-based model structures. Techniques designed to incorporate the predictive models into on-line monitoring and control of batch product quality were also examined. The proposed control approach can be viewed as shrinking-horizon model-predictive control based on empirical models. The effectiveness of the proposed prediction and control methods are illustrated by using an industrially relevant simulated polymerization example.

Introduction

Overview of the problem

Despite having a lower production volume when compared to continuous processing, batch processing offers some important advantages, like easy scale-up from laboratory procedures and a high degree of operating flexibility. Due to these advantages, batch processing is used extensively in the production of low-volume, high value-added products, including polymers, pharmaceuticals, and specialty chemicals. The "quality" of the product from such processes is often a subjective quantity, depending on whether one is concerned with ease of production, sales success, or the end-use performance of the product. In this article, we define product quality as those engineering variables of interest that are most closely related to the end-use characteristics and economic success of the product. For example, product quality may be composed of number average molecular weight and polydispersity for a polymerization reactor (Schork et al., 1993; Russell et al., 1998b) or kappa number for a pulp digester (Lee and Datta, 1994). Over the last two decades, increased global competition has resulted in higher demand for the production of uniform quality products from batch and semibatch processes. In turn, industry is beginning to respond to the economic opportunity associated with improved quality control of such processes.

At this point, we begin the discussion by defining the system under consideration shown in Figure 1. First, there is the batch or semibatch process itself, with initial (feed) conditions x_0. It is assumed that the initial conditions are a function of unmeasurable feed disturbances (ω_f) that may vary from batch to batch. The control objective is to manipulate the process inputs (u) to "steer" the process as well as overcome any disturbances (ω) so that the batch evolves in such a manner as to achieve the desired final product quality (q). Almost invariably, the quality variables of interest are not measurable on-line, rendering direct feedback controller design infeasible.

Traditionally, statistical process control (SPC) has been the most extensively used approach when only off-line quality measurements are available (Ott and Schilling, 1990). In SPC, the final product quality is measured at the end of the batch and is compared to a statistical base of "normal operation." If the quality measurements show a statistically significant deviation from the normal operating base, the operator or engineer may systematically look for a cause or adjust the operating condition for the next batch in an attempt to re-

Correspondence concerning this article should be addressed to J. H. Lee.

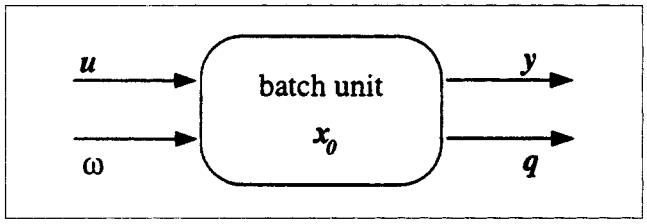

Figure 1. Batch process.

duce the variation. Hence, SPC in the traditional sense is an off-line technique (we add the qualifier "traditional" here since the concept has recently been expanded by MacGregor and coworkers to include on-line measurements, as we shall discuss later). Although SPC may potentially be effective for detecting abnormal situations (e.g., faults, failures) or dealing with persistent upsets that last through several batches, the exclusive use of off-line measurements severely limits the controller's ability to respond to batch-to-batch variations.

To take advantage of on-line measurement information, engineers often attempt to control some measurable secondary process variables (y), which are usually difficult to relate directly to the final product quality. Typical secondary measurements include easily measured quantities such as temperatures, pressures, and flows. The most common of these secondary variable control approaches involves controlling the measurable process variables to follow reference trajectories that are determined *a priori* from an off-line model or the historical data of "good" batch runs (Russell et al., 1998b). This can be effective in rejecting some of the disturbances (e.g., heat-transfer variations, valve errors). However, for systems that experience significant batch-to-batch variations in feed conditions and/or reaction parameters (for example, mass-transfer coefficients or kinetic parameters), this strategy can result in large quality variations, even with perfect tracking controls (Russell et al., 1998b). In many cases, the performance of the tracking control system itself may vary from batch to batch, adding to the problem. Such an approach frequently produces off-spec product that must be stored at some expense and blended with other batches or else becomes waste. Given the shortcomings of the current approaches, a significant incentive exists for developing more flexible and effective quality-control techniques.

Ideally, one would like a more flexible control strategy that still retains the use of the information contained in the on-line measurements. Consider the relationship between the various deviations from the nominal trajectories of the on-line measurable variables and the final product quality for past batch runs, as depicted in Figure 2. For most systems, variations in the trajectories of the on-line secondary measurements will lead to variations in the final product quality. On the other hand, the change in any one variable by itself may be difficult for an operator to detect or evaluate its significance (MacGregor et al., 1996). Such measurement trajectories represent the "footprint" of the process because they contain a great deal of information about how the batch is progressing in terms of the development of the final product quality. The remainder of this article is devoted to examining techniques designed to capture this information in a model that is capable of predicting final product quality on the basis

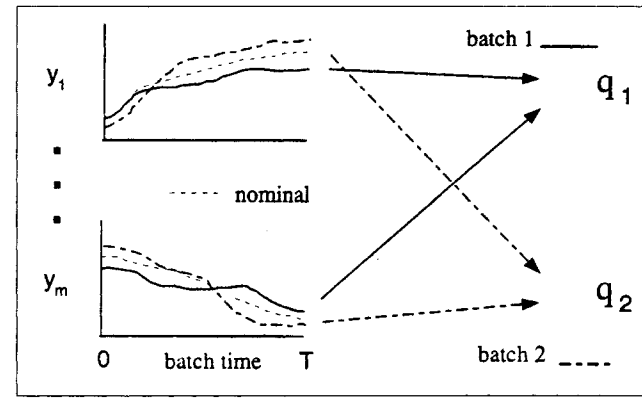

Figure 2. Relationship between measurement trajectories and final quality.

of on-line measurements. Once an on-line quality prediction model is available, it may then be used to monitor the development of the final product quality and calculate control moves designed to achieve the given quality targets. In the next sections, we examine some techniques for realizing these goals from a data-based, empirical regression model perspective.

Perspective on the current literature

Before formulating the general data-based quality prediction and control approach that is to be the basis of this article, we give a brief overview and perspective of some of the currently available literature related to the problem under consideration.

MacGregor and coworkers in several related articles (Kourti and MacGregor, 1995; Kresta et al., 1991; Nomikos and MacGregor, 1994, 1995a,b; MacGregor et al., 1994) introduced a new SPC technique for batch and semibatch processes based on the concept of multiway PCA. The basic premise of their approach is the same as in the traditional SPC, but the statistical monitoring is done using on-line process measurements rather than off-line quality measurements. With this technique, an abnormal batch can be detected quickly (at the onset of abnormality), providing an opportunity to stop the batch or make compensatory adjustments. For many batch systems, however, the dividing line between normal batches and abnormal batches is not clear-cut. Quality variations even among the normal batches can be quite significant. In addition, some of the abnormal batches may be saved by making appropriate compensations to the batch condition. This motivates development of an inferential quality control scheme that controls quality variables on an active basis.

In order to build an empirical regression model for on-line quality prediction, certain problems associated with the nature of the batch data must be addressed. For instance, conditioning problems related to the extremely large number of highly correlated variables involved in building data-based regression models are a primary concern. Most of the related literature in this field is centered around multivariate statistical modeling techniques like principal component regression (PCR) and partial least squares (PLS) (Geladi and Kowalski,

1986; MacGregor et al., 1991), which give certain advantages in terms of data reduction and numerical conditioning. In this article, we explore the use of a single regression model to make *recursive* predictions of the final product quality at any point during the batch using *only* the on-line measurements available up to that point. Unlike previous researchers who have relied on various *ad hoc* assumptions, we will use the Kalman filter theory to produce statistically optimal recursive quality predictions.

In contrast to the monitoring, the use of statistical modeling techniques in automatic batch process quality control is only just beginning to be explored. Most of the work in early articles was oriented more toward the use of PLS models for controlling continuous processes around certain operating conditions (Budman et al., 1991 and 1992; Ye et al., 1993; Kaspar and Ray, 1992, 1993a,b; Qin and McAvoy, 1996; Larimore, 1997; Cheng and McAvoy, 1997; Ku et al., 1995). In one of the earliest significant works, Piovoso and Kosanovich developed a steady-state feedback controller design that used a static PCA/PCR model to manipulate process inputs in an attempt to remain in the region of statistically acceptable operation (Piovoso and Kosanovich, 1994; Piovoso et al., 1992). In a related work, Cheng and McAvoy (1996) extended this approach by reformulating the control calculation as an optimization problem. With this improved structure, robustness to disturbances and model errors can be improved and process constraints can be handled. It is felt that this approach offers some interesting possibilities that were not fully examined. Thus, several of the ideas in these articles serve as inspirations for some for the control developments to be presented in this work.

Dong, McAvoy, and Zafiriou have taken a different approach aimed more at using nonlinear neural network PLS models for batch-to-batch optimization of input profiles while explicitly considering process constraints (Dong et al., 1996). From an on-line control perspective, the major disadvantage of their approach is the practice of only updating the trajectories between batches. The resulting off-line optimal input trajectories can be quite sensitive to process disturbances, especially those affecting the initial conditions of the batch. The control approach to be proposed in this article can be interpreted as being similar, except that the optimal profiles are updated throughout the batch. This increased frequency of control move calculation will enable response to any process disturbances that may be detected by on-line measurements.

In a recent work, Yabuki and MacGregor (1997) used on-line and infrequent off-line measurements at some midcourse point to predict the final product quality using PLS models. If the predicted quality deviations were outside a statistically defined "no control" deadzone, a model was then used to calculate a semibatch control move designed to bring the batch back under statistical control. The approach presented in Yabuki's article can easily be extended to the pure batch case and to cases where infrequent off-line measurements are not available. In the present article, we will generalize the approach and offer more insight into the problem. A more rigorous approach to calculating the control moves will also be taken here, one that is based on an optimization problem similar to the shrinking-horizon model predictive control (MPC) of Joseph and Hanratty (1993) where constraints are specifically considered.

Finally, Joseph and Hanratty have developed an approach to batch quality control that relies on neural network models (Joseph and Hanratty, 1993). Essentially, the concept can be described as shrinking-horizon MPC and, except for the choice of model structures, is similar to the control approach to be proposed here. However, the initial conditions are considered to be known or measurable in Joseph's article, whereas they will be more realistically modeled as unmeasurable disturbances in our approach. In Joseph and Hanratty (1993), the authors chose to handle the missing future measurements in an *ad hoc* fashion by making assumptions concerning the missing data, which can potentially lead to poor predictions. In the approach to be proposed in this article, the problem of the missing future measurements will be handled in a statistically optimal manner.

Intended contribution

In this article, a general data-based approach to obtaining on-line predictions of final product quality is developed. The general formulation allows for the extension of the previous literature results to the problems of (1) prediction of product quality under disturbance and control effects, (2) recursive quality prediction, and (3) on-line data-based control of product quality. Finally, an application of the proposed methods to the industrially relevant nylon 6,6-polymerization process is be examined.

Organization

The general data-based approach to on-line quality prediction is formulated in the following section. In that section, a general least-squares regression approach is first considered along with the proper choice of data for constructing the model. Subsequently, an equivalent statistical approach to the prediction problem is presented and a recursive quality-prediction algorithm based on Kalman filter theory is developed. Biased regression techniques, including multivariate statistical modeling methods, are incorporated to handle problems associated with the batch data. In the third section, optimal control techniques for using the data-based prediction models in on-line quality control are presented. The techniques presented in the previous sections are applied to an industrially relevant polymerization example in the fourth section, and some concluding remarks are made in the fifth section.

On-Line Prediction of Product Quality

In this section, the general data-based approach to on-line quality prediction is formulated for the conditions outlined in the Introduction. After the underlying system model is presented, the motivations for the data-based approach are developed based on the shortcomings of the fundamental-model-based approach. The general quality prediction model form based on least-squares regression is then examined and the data required to realize the model form are discussed. Biased regression techniques like PCR and PLS are then proposed as methods to achieve the data-based model form when faced with certain problems associated with the size and correlated nature of batch data. Finally, an equivalent statistical interpretation of the quality prediction problem is

presented, which allows for an easy formulation of a recursive predictor and on-line monitoring of model validity.

General framework

To lay the basis for the discussion, it is assumed that the batch or semibatch process under consideration can be described by the following nonlinear ODE system

$$\frac{dx}{dt} = f_x(x, u, \omega) \qquad y = f_y(x, \upsilon), \tag{1}$$

with the initial condition $x_0 = x(t = 0)$. The vector x represents the process state, the vector u represents the manipulated inputs, the vectors ω and υ represent state and output disturbances, and the vector y represents the measured process output variables. As an example, the states of a polymerization reactor may include temperature, concentrations of the chemical components, and parameters related to the molecular weight distribution. In addition to the state-space description of Eq. 1, it is further assumed that the final product quality (q) is a function of the final state of the system [$x_{t_f} = x(t = t_f)$] as given by the following relation:

$$q = f_{qx}(x_{t_f}). \tag{2}$$

For a polymer reactor, product quality typically includes hard-to-measure quantities such as number average molecular weight, polydispersity, or concentrations of key functional groups (Schork et al., 1993; Russell et al., 1998b; Schuler and Suzhen, 1985).

According to the structure of Figure 1, the process outputs may be the measurable process variables (temperatures, pressures, etc.), while the inputs may correspond to the actuator positions (e.g., valve positions). More commonly, low-level loops may exist. In this case, the inputs may be considered to be the set points to the low-level loops. The resulting actuator positions or associated variables (e.g., temperature or pressure of heating/cooling medium in a temperature loop) may be included as outputs, if they provide useful information. Alternatively, one can use the process variables that are controlled by the low-level loops as inputs if the performance of these loops is sufficiently good.

Fundamental model-based approach

One of the prevailing trends in current batch chemical process control research is to develop a fundamental, state-space model of the system and use this as the basis of a model-based approach to predict product quality. Suppose a model of the following form is given for system in Eq. 1:

$$\frac{dz}{dt} = f_z(z, u, \omega_z) \qquad y = f_y(z, \upsilon). \tag{3}$$

Besides the system model, the quality model is specified as

$$q = f_{qz}(z_{t_f}).$$

Once these descriptions are available, the next step in this process is to construct a state estimator based on the nonlinear model. Typically, one first assumes that z_0, ω_z, and υ are random variables with certain statistical distributions (which are chosen based on some *a priori* process knowledge and further "tuned" later). Basically, the uncertainty in z_0 reflects feed disturbances, ω_z characterizes model errors and unmeasured disturbances, and υ represents measurement errors. From this information, one could construct a nonlinear observer (Jazwinski, 1970; Robertson et al., 1996; Gelb, 1974) of the following form:

$$\hat{z}_{i|i} = f_{\text{obs}}(y_1, \ldots, u_0, \ldots, u_{i-1}, \hat{z}_{0|0}, \mathcal{P}_{0|0}, R_\omega, R_\upsilon), \tag{4}$$

which is based on the process model, the known inputs u and the measurements of y up to time i, the estimated initial condition $\hat{z}_{0|0}$, and the statistical description of the random variables. Throughout the remainder of the discussion, the notation $\hat{\zeta}_{i|j}$ will denote the estimate of the random variable ζ made at time i given all of the information up to time j.

The prediction of the final state $\hat{z}_{t_f|i}$ (where t_f is the total number of samples within a single batch) may be obtained by integrating the model from the current state estimate $\hat{x}_{i|i}$ with known inputs and the best estimates of the future noise terms based on the assumed statistical description (Jazwinski, 1970).

$$\hat{z}_{t_f|i} = f_{\text{pred}}\left(y_1, \ldots, y_i, u_0, \ldots, u_{t_f-1}, \hat{z}_{0|0}, \mathcal{P}_{0|0}, R_w, R_\upsilon\right). \tag{5}$$

In most cases, one is likely to attempt to reduce the computational difficulties by using an extended Kalman filter (EKF) (Jazwinski, 1970) as the nonlinear observer in Eq. 4. Subsequently, predictions of the final product quality may be obtained from the predictions of the final state by substituting into the quality model given by

$$\hat{q}_{t_f|i} = f_{qz}\left(\hat{z}_{t_f|i}\right). \tag{6}$$

The objective of the model-based approach is to use the process knowledge contained in the model to simplify the prediction and improve its accuracy. However, there are many difficulties incurred in developing the model, estimating various parameters, and modeling the random disturbance components that undermine this goal to a large extent. Even when the state-space model can be developed, it is often very difficult to relate effectively the product quality to the final state of the system. This is because product quality may not always be completely specified by the states of a finite-dimensional, parsimonious model description (Schuler and Suzhen, 1985) (since the model state z may have a significant lower dimension than that of the actual system state x). For instance, the quality may depend on additional information such as the molecular weight distribution for a polymerization reactor (Russell et al., 1998a). In addition, formulating the observer/predictor and solving the computationally demanding estimation problem on-line significantly adds to the complexity of this approach for a realistic system.

Data-based approach: Least-squares framework

In this subsection, an alternative approach to predicting the product quality variables is developed that is based completely on regression models obtained from the process data. The premise is to replace the fundamental model with the process knowledge captured in the historical record of process data. Thus, a quality predictor analogous to combining the final state predictor in Eq. 5 with the quality model in Eq. 6 is built directly from the data without the difficult intermediate step of developing a fundamental model and state observer.

Formulating the Regression Problem. As a first step to realizing the stated objective, consider the following general model for the product quality variables that can only be measured after a lengthy off-line analysis: first, assume that $\hat{z}_{0|0}$, $\mathcal{P}_0$, R_ω, and R_υ are constant. Second, combine Eqs. 5 and 6 to form

$$\hat{q}_{t_f|i} = f_q(y_1, \ldots, y_i, u_0, \ldots, u_{t_f-1}). \tag{7}$$

Note that the observer takes the preceding form irrespective of the state dimension. Infrequent off-line measurements may also be included in this structure provided they are available before the end of the batch. To simplify the notation, let the vector of on-line process measurements up to time i be $\mathcal{Y}_i = [y_1^T, \ldots, y_i^T]^T$, while $\mathcal{U}_i$ is the vector of the manipulated inputs up to time i: $\mathcal{U}_i = [u_0^T, \ldots, u_{i-1}^T]^T$. For the remainder of the discussion, the absence of a subscript on $\mathcal{Y}$ or $\mathcal{U}$ will imply the full data vector at $i = t_f$. Thus, Eq. 7 can be cleanly restated as

$$\hat{q}_{t_f|i} = f_q(\mathcal{Y}_i, \mathcal{U}). \tag{8}$$

One can linearize the general model in Eq. 8 with respect to some "nominal" (mean or reference) measurement trajectory $y_1^r, \ldots, y_{t_f}^r$ and corresponding nominal input trajectory $u_0^r, \ldots, u_{t_f-1}^r$ in order to arrive at the following linearized model:

$$\hat{q}'_{t_f|i} = A\,\mathcal{Y}'_i + B\,\mathcal{U}', \tag{9}$$

where $\mathfrak{M}' = \mathfrak{M} - \mathfrak{M}^r$ for vector $\mathfrak{M}$, and q^r is the desired target quality. In many cases, subtracting the time-varying reference trajectory from the raw trajectories removes the majority of the process nonlinearity and allows linear modeling methods to perform well on the resulting deviation variables. Therefore, most of the remaining discussion will center on the linearized model; however, the primes and the t_f on q'_{t_f} will be suppressed for clarity ($q'_{t_f} = q$ and $\hat{q}'_{t_f|i} = \hat{q}_i$).

In this research, the objective will be to utilize data-based modeling techniques in order to develop a relation between the various on-line measurements available during the batch, the known manipulated input variables, and the unmeasurable final-quality variables as in Eq. 9. This model will then be used to predict the final quality and manipulate the inputs to achieve target quality values.

Least-Squares Regression. The first question we must now address concerns how the general regression model in Eq. 9 may be obtained. At this point, for the sake of clarity we will restrict our attention to the case where only on-line measurements are considered; this corresponds to the case where the input moves are kept at the nominal values. As is described later, the effects of the deterministic manipulated input moves can easily be incorporated at a later point using the framework of Eq. 9.

At sample time i ($1 \le i \le t_f$), we desire to build a linear predictor for final quality using the available measurements:

$$\hat{q}_i = \hat{\Theta}_i \mathcal{Y}_i \tag{10}$$

where $\hat{\Theta}_i$ is the model coefficient matrix at time i. Now, suppose we are given the appropriate quality and on-line measurement data from N past batches, $q(1), \ldots, q(N)$ and $\mathcal{Y}_i(1), \ldots, \mathcal{Y}_i(N)$, where index (j) refers to the data from batch number j. Then a reasonable choice for $\hat{\Theta}_i$ is the argument minimizing the following objective function:

$$\min_{\Theta_i} \sum_{j=1}^{N} \| q(j) - \hat{\Theta}_i \mathcal{Y}_i(j) \|_2^2. \tag{11}$$

The solution can be obtained from the equivalent least-squares problem for $Q = \hat{\Theta}_i Y_i + E$, where

$$Y_i = [\mathcal{Y}_i(1) \cdots \mathcal{Y}_i(N)] \qquad Q = [q(1) \cdots q(N)], \tag{12}$$

and the Frobenieus norm of the residual matrix, $\| E \|_F$, is to be minimized. Y_i is of dimension $(i \cdot n_y) \times N$, where n_y is the number of measurements at each time instant, while Q is of dimension $n_q \times N$, where n_q is the number of final-quality variables. The least-squares solution may be expressed as

$$\hat{\Theta}_i^{LS} = Q Y_i^T (Y_i Y_i^T)^{-1}, \tag{13}$$

with the prediction model now becoming $\hat{q}_i = \hat{\Theta}_i^{LS} \mathcal{Y}_i$. The inverse of the matrix $Y_i Y_i^T$ may be "ill-behaved," an issue that is addressed next.

Handling Colinearity Through Biased Regression. The recent development of more powerful, less expensive computers and measurement devices has led to a wealth of information stored in databases for most industrial processes. Typically, measurements of several secondary variables are taken at evenly spaced time intervals throughout the batch. Rarely do each of these measurements provide independent information. It is much more likely that the process is only free to vary in an underlying small-dimensional space, resulting in a large degree of correlation in the observations. Constructing the empirical model in Eq. 9 via traditional regression techniques such as ordinary least squares, as in Eq. 13, is not appropriate in this case because of the serious conditioning problems resulting from the large number and colinearity of the observations. The numerical problems (e.g., $Y_i Y_i^T$ becoming singular) caused by this ill-conditioning result in a high degree of sensitivity of the unbiased estimate $\hat{\Theta}$ to the data, and thus a large variance in $\hat{\Theta}$.

We can overcome both the dimensional and conditioning problems by accepting some bias in our estimate $\hat{\Theta}$ in order to reduce the variance. This is accomplished using the gen-

eral class of biased regression techniques (e.g., PCR and PLS) to estimate Θ [excellent overviews of these techniques are given in Geladi and Kowalski (1986), MacGregor et al. (1991), and Hoskuldsson (1988)]. The effectiveness of these methods comes from their ability to utilize the correlation contained in the data record to reduce the dimensionality of the problem. This reduction in dimensionality is achieved by projecting the original variables onto lower-dimensional subspaces:

$$\begin{aligned} \mathcal{Y}_i &\xrightarrow{P_i} \underline{\mathcal{Y}_i} & \underline{\mathcal{Y}_i} &= P_i \mathcal{Y}_i \\ q &\xrightarrow{R} \underline{q} & \underline{q} &= Rq. \end{aligned} \tag{14}$$

In this approach, the $n_{yr} \times (i \cdot n_y)$ and $n_{qr} \times n_q$ dimensional orthogonal projection matrices P_i and R are built from the correlation structure of the data according to the particular multivariate statistical method used (please see the referred articles for details). Thus, the regression may be formulated in the well-conditioned reduced space according to

$$\underline{\hat{\Theta}_i} = \underline{Q Y_i^T} \left(\underline{Y_i Y_i^T} \right)^{-1}, \tag{15}$$

with the following prediction equation:

$$\hat{q}_i = (R^T \underline{\hat{\Theta}_i} P_i) \mathcal{Y}_i = \hat{\Theta}_i^{BR} \mathcal{Y}_i. \tag{16}$$

In the preceding, superscript *BR* stands for *biased regression.*

Choosing the Number of Reduced Variables. Ideally, the number of reduced variables (e.g., latent variables for PLS or principal components for PCR) used in the model would be equal to the number of modes of variation present in the process. In practice, the inherent nonlinearity of many processes and the presence of noise may cause a few more reduced variables to be required to represent the information present in the original data matrix. In general, no one method for choosing the number of reduced variables has achieved universal acceptance. Often, it has been advocated that one should choose the number of reduced variables to be included in the model based on some criterion that expresses how well the output quality data (Q) are explained. For example, the number of latent variables is often chosen by minimizing the sum of squares for the output (quality) prediction error (SPE_q).

$$SPE_q = \sum_{j=1}^{N_{\text{test}}} [q(j) - \hat{q}(j)]^T [q(j) - \hat{q}(j)] \tag{17}$$

of an *independent* model-testing data set. Using these criteria, it is possible that one could explain most of the off-line quality variance using only a few reduced variables, while the variance of the corresponding explained on-line measurement data (Y_i) is relatively low. For the prediction, monitoring, and control purposes of this article, we are interested in explaining the maximum amount of variance in both Y_i and Q. Thus, we select the number of reduced variables required to describe the desired amount of the measurement data variance and output quality data variance *separately* and independently. The number of reduced variables generally included in the model is chosen as the maximum needed for adequately describing either data set. A careful analysis of the percent variance explained by each reduced variable is necessary to ensure that overfitting has not occurred, which would result in an increased sensitivity to noise.

Multiple-Model vs. Single-Model Approach. Using the results of the previous section, regression models of the form in Eq. 9 may now be constructed from historical data. Such a model allows for prediction of quality once the necessary process measurements are available ($\mathcal{Y}_i$ is complete). However, practical process monitoring and control applications normally require that predictions be made at several midpoints in the batch *before* all of the measurements can be available (ideally, a recursive approach where predictions can be made at each sample time on the basis of a single model would be desired). This presents a problem concerning how to handle the measurements that are not yet available when predictions are required.

Basically, there are two approaches one could take to address this problem. The first is a multiple-model approach where one builds a separate model for each instant where the prediction is desired, based only on the data available up to that point. This approach is simple in that it avoids problems with the unavailable future data, but it may become tedious to build a separate model for each of several desired prediction points.

Alternatively, a single model may be used to obtain quality predictions provided some assumptions are made about the unavailable future measurements. The basic idea is that one could build a single model at the final sample time $\hat{q}_{t_f} = \hat{\Theta}_{t_f} \mathcal{Y}_{t_f}$ and replace the full measurement vector $\mathcal{Y}_{t_f}$ with its prediction based on the currently available measurements $\hat{\mathcal{Y}}_{t_f|i}$, resulting in $\hat{q}_i = \hat{\Theta}_{t_f} \hat{\mathcal{Y}}_{t_f|i}$. One could assume that the future measurements are simply given by their nominal values, as in Nomikos and MacGregor (1995b) and Joseph and Hanratty (1993), or that the error in the current measurement will remain constant throughout the remainder of the batch, as in Nomikos and MacGregor (1994). Because both of these assumptions may lead to inaccurate predictions, Nomikos and MacGregor have advocated a more consistent approach using PLS models (Nomikos and MacGregor, 1995a). Basically, the future measurements are constrained in that article to be consistent with the previously obtained observations and the correlation captured in the PLS model. This approach is not truly recursive, however, but requires a full batch processing of all the available data at each step. In addition, the estimate can be extremely sensitive to measurement noise and outliers during the early phase of a batch. In the remainder of this section, we develop a more general, fully recursive approach to batch quality prediction that is applicable for any regression model of the chosen form.

Incorporating Deterministic Inputs. For pure monitoring purposes, there need not be a distinction between inputs and outputs (both can be treated simply as observed variables). However, since we are interested in using the model for control as well, the distinction becomes necessary. Inputs here are those that can be manipulated for control purposes, perhaps within some constraints.

Having developed the predictor for the measurement-only case, we are now in a position to include the effects of deter-

ministic input moves contained in the $(n_u \cdot t_f) \times N$ matrix U composed of the n_u manipulated input moves at t_f intervals for the past N batches $(\mathfrak{U}(1), \ldots, \mathfrak{U}(N))$. This can be accomplished in a straightforward manner by first augmenting $\mathfrak{Y}_i$ with $\mathfrak{U}$ to form the $(i \cdot n_y + n_u \cdot t_f) \times 1$ augmented vector $\mathfrak{Z}_i = [\mathfrak{Y}_i^T \ \mathfrak{U}^T]^T$. (Similarly, the historical data matrix Z_i is formed by augmenting Y_i with U.) Just as was done for the measurement-only case, one can construct a prediction model of the form:

$$\hat{q}_i = \hat{\Theta}_i^{LS} \begin{bmatrix} \mathfrak{Y}_i \\ \mathfrak{U} \end{bmatrix} = \hat{\Theta}_i^{LS} \mathfrak{Z}_i, \tag{18}$$

where

$$\hat{\Theta}_i^{LS} = Q_i Z_i^T (Z_i Z_i^T)^{-1}. \tag{19}$$

One can also perform biased regression just as before by projecting the augmented vector $\mathfrak{Z}_i$ down to a reduced-dimensional space:

$$\begin{array}{lll} \mathfrak{Z}_i \xrightarrow{P_i} \underline{\mathfrak{Z}}_i & \underline{\mathfrak{Z}}_i = P_i \mathfrak{Z}_i \\ q \xrightarrow{R} \underline{q} & \underline{q} = Rq. \end{array} \tag{20}$$

The corresponding least-squares solution is therefore:

$$\underline{\Theta}_i^{BR} = \underline{Q}_i \underline{Z}_i^T (\underline{Z}_i \underline{Z}_i^T)^{-1} \tag{21}$$

$$\hat{\Theta}_i^{BR} = R^T \underline{\hat{\Theta}}_i^{BR} P_i \tag{22}$$

In some cases, one may wish to use separate projections for $\mathfrak{Y}_i$ and $\mathfrak{U}$:

$$\begin{array}{lll} \mathfrak{Y}_i \xrightarrow{S_i} \underline{\mathfrak{Y}}_i & \underline{\mathfrak{Y}}_i = S_i \mathfrak{Y}_i \\ \mathfrak{U} \xrightarrow{V} \underline{\mathfrak{U}} & \underline{\mathfrak{U}} = V \mathfrak{U} \end{array} \tag{23}$$

Using the separate projections, the augmented vector may be formulated as

$$\underline{\mathfrak{Z}}_i = \begin{bmatrix} Si & 0 \\ 0 & V \end{bmatrix} \mathfrak{Z}_i = P_i \mathfrak{Z}_i \qquad P_i = \begin{bmatrix} S_i & 0 \\ 0 & V \end{bmatrix}. \tag{24}$$

Choice of the Model-Building Data Set. An important topic to be addressed on the subject of data-based prediction models concerns the choice of the data set used to build the models found in Eqs. 9 and 13. It is important to recognize that the predictive capabilities of data-based models are heavily dependent on the nature of the data supplied for building the model.

One of the goals in this article is to produce empirical models that allow effective control moves to be calculated for dealing with common process upsets. Therefore, the data used to build the model must contain batches in which these disturbances were present and control moves were made to compensate for the disturbance effects. This will allow the correlation that exists between disturbances, input moves, measured process outputs, and final product quality to be modeled. Thus, the "region of normal operation" for these control-relevant models has a somewhat more expanded meaning than for those of pure monitoring purposes. It should be defined to include any batch that can be "saved" (i.e., steered to the desired final state) by making compensatory adjustments on the manipulated variables.

By the same token, care must be taken that the input moves represented in the data are sufficiently rich to allow meaningful identification of the input-effect model (i.e., B in Eq. 9). For some systems, designed input experiments may be required, with input moves made solely for identification purposes.

Potentially, incorporating large disturbances into the normal operating data set can widen the range of operation sufficiently for process nonlinearities to become significant. In this case, a locally weighted (Wise and Gallagher, 1997) or nonlinear (Wise and Gallagher, 1997; Qin and McAvoy, 1992) data-based model may be required in order to maintain good predictive capability.

Data-based approach: Statistical viewpoint

In this subsection, we reexamine the regression problem from a statistical perspective. The statistical perspective offers many advantages including an intuitive interpretation, the ability to place meaningful confidence intervals on the resulting quality predictions, and a recursive formulation on the basis of the Kalman filter.

Equivalent Statistical Estimator. To begin, we restrict our attention to the case where $i = t_f$ (thus $\mathfrak{Y} = \mathfrak{Y}_{t_f}$). Let us assume that both $\mathfrak{Y}$, $\mathfrak{U}$, and q are zero-mean Gaussian variables with the covariance matrix [when $\mathfrak{U}$ is viewed as a deterministic vector, the covariance matrices $\mathcal{P}_{yu}$, etc., are to be interpreted as $\lim_{n \to \infty} \Sigma_{i=1}^{n} \mathfrak{Y}(i) \mathfrak{U}^T(i)$, etc.; quasi-stationarity is assumed here] of

$$E \begin{Bmatrix} \mathfrak{Y}\mathfrak{Y}^T & \mathfrak{Y}\mathfrak{U}^T & \mathfrak{Y}^T q \\ \mathfrak{U}\mathfrak{Y}^T & \mathfrak{U}\mathfrak{U}^T & \mathfrak{U}^T q \\ q\mathfrak{Y}^T & q\mathfrak{U}^T & qq^T \end{Bmatrix} = \begin{bmatrix} \mathcal{P}_y & \mathcal{P}_{yu} & \mathcal{P}_{yq} \\ \mathcal{P}_{uy} & \mathcal{P}_u & \mathcal{P}_{uq} \\ \mathcal{P}_{qy} & \mathcal{P}_{qu} & \mathcal{P}_q \end{bmatrix} \tag{25}$$

For the situation under consideration, the statistically optimal predictor of q is simply the conditional expectation of q, given the measurements and inputs:

$$\hat{q} = E\{q \mid \mathfrak{Y}, \mathfrak{U}\} = [\mathcal{P}_{qy} \ \mathcal{P}_{qu}] \begin{bmatrix} \mathcal{P}_y & \mathcal{P}_{yu} \\ \mathcal{P}_{uy} & \mathcal{P}_u \end{bmatrix}^{-1} \begin{bmatrix} \mathfrak{Y} \\ \mathfrak{U} \end{bmatrix}. \tag{26}$$

Under normal circumstances, the covariance matrices in these equations are not known. Suppose we calculate the sample covariance matrices using the historical measurement data from the past N batches according to

$$\begin{bmatrix} \hat{\mathcal{P}}_y & \hat{\mathcal{P}}_{yu} \\ \hat{\mathcal{P}}_{uy} & \hat{\mathcal{P}}_u \end{bmatrix} = \frac{1}{N}\sum_{j=1}^{N}\begin{bmatrix} \mathcal{Y}(j) \\ \mathcal{U}(j) \end{bmatrix}[\mathcal{Y}^T(j)\mathcal{U}^T(j)]$$

$$= \frac{1}{N}\begin{bmatrix} Y \\ U \end{bmatrix}[Y^T\ U^T] \quad (27)$$

$$\begin{bmatrix} \hat{\mathcal{P}}_{qy} & \hat{\mathcal{P}}_{qu} \end{bmatrix} \approx \frac{1}{N}\sum_{j=1}^{N} q(j)[\mathcal{Y}^T(j)\ \mathcal{U}^T(j)] = \frac{1}{N}Q[Y^T\ U^T]. \quad (28)$$

Note. How the nominal trajectories were determined affects the degree of freedom in the preceding calculation of sample covariances. If the trajectories were determined from the same data used to compute the covariances, the degree of freedom is $N-1$ rather than N.

By substitution, the optimal prediction of q is given by

$$\hat{q} = Q[Y^T\ U^T]\left(\begin{bmatrix} Y \\ U \end{bmatrix}[Y^T\ U^T]\right)^{-1}\begin{bmatrix} \mathcal{Y} \\ \mathcal{U} \end{bmatrix} = \hat{\Theta}^{LS}\begin{bmatrix} \mathcal{Y} \\ \mathcal{U} \end{bmatrix}. \quad (29)$$

In this case, one simply recovers the least-squares solution of Eq. 13.

Since the covariance matrices are estimated using the available sample, the potential for the strong adverse effect of the estimation error on the quality prediction must be guarded against. The covariance approximation error can have a strong effect on the estimation when the correlations among the elements of $\mathcal{Y}$ and $\mathcal{U}$ are strong, and hence the matrix ZZ^T (where $Z=[Y\ U]$) is ill-conditioned. Given the equivalence with the least squares, the problem of colinearity can be handled in the same manner as before, that is, by projecting the $\mathcal{Y}$, $\mathcal{U}$, and q data down to the appropriate subspaces:

$$\begin{bmatrix} \mathcal{Y} \\ \mathcal{U} \end{bmatrix} \xrightarrow{P} \underline{Z} \qquad q \xrightarrow{R} \underline{q}. \quad (30)$$

The optimal predictor for q may be formulated on the basis of the conditional expectation in the reduced space just as before:

$$\hat{q} = R^T\hat{\mathcal{P}}_{\underline{qz}}\hat{\mathcal{P}}_{\underline{z}}^{-1}P\begin{bmatrix} \mathcal{Y} \\ \mathcal{U} \end{bmatrix}. \quad (31)$$

The sample covariance matrices may again be computed from the projected data for the past N batches according to

$$\hat{\mathcal{P}}_{\underline{qz}} \approx \frac{1}{N}\sum_{j=1}^{N}\underline{q}(j)\underline{Z}^T(j) = \frac{1}{N}\underline{QZ}^T \quad (32)$$

$$\hat{\mathcal{P}}_{\underline{z}} \approx \frac{1}{N}\sum_{j=1}^{N}\underline{Z}(j)\underline{Z}^T(j) = \frac{1}{N}\underline{ZZ}^T. \quad (33)$$

Note that this formulation is equivalent to the least-squares estimator obtained through biased regression ($\hat{\Theta}^{BR}$) in Eqs. 15 and 16.

Recursive Least-Squares Prediction through the Kalman Filter. In this section, we examine the objective of obtaining statistically optimal *recursive* predictions of product quality that are updated with each incoming measurement on the basis of a single empirical model. We begin by defining the "state" variable ζ_i as *all of the measurements within the batch* and the final quality ($[\mathcal{Y}_{t_f}^T\ q^T]^T$) with the future input moves for $i,\ldots,t_f-1$ held at their respective set points ($u_i = u_{i+1} = \cdots = u_{t_f-1} = 0$). Based on this, one can write the time evolution of the state variable as

$$\zeta_{i+1} = \zeta_i + B_i u_i, \quad (34)$$

where the effects of the input u_i are incorporated through the matrix B_i, which is the ith column block (of size n_u) of

$$\begin{bmatrix} \mathcal{P}_{yu}\mathcal{P}_u^{-1} \\ \mathcal{P}_{qu}\mathcal{P}_u^{-1} \end{bmatrix}.$$

The measurement at the current time is simply the corresponding element of the state variable:

$$y_i = \underbrace{[[0\cdots 0\ I_{n_y}\ 0\cdots 0][0_{n_q}]]}_{\mathcal{C}_i}\zeta_i, \quad (35)$$

where I_{n_y} is the identity matrix of size n_y located in the appropriate position corresponding to the current measurement.

Continuing with the statistical framework presented in the last section for jointly Gaussian $\mathcal{Y}$ and q, the statistically optimal linear *recursive* predictor of q may again be developed on the basis of the conditional expectation. This recursive quality predictor takes the form of a Kalman filter with initial state ζ_0, which is $[\mathcal{Y}^T\ q^T]^T$ with the effect of the deterministic inputs removed. Hence, it is zero-mean Gaussian with covariance matrix

$$\mathcal{P}_0 = E\{\zeta_0\zeta_0^T\} = \begin{bmatrix} \mathcal{P}_y & \mathcal{P}_{yq} \\ \mathcal{P}_{qy} & \mathcal{P}_q \end{bmatrix} - \begin{bmatrix} \mathcal{P}_{yu} \\ \mathcal{P}_{qu} \end{bmatrix}\mathcal{P}_u^{-1}[\mathcal{P}_{uy}\ \mathcal{P}_{uq}]. \quad (36)$$

The Kalman filter equations are given by

$$\hat{\zeta}_{i+1} = \hat{\zeta}_i + B_i u_i + \mathcal{K}_{i+1}[y_{i+1} - \mathcal{C}_{i+1}(\hat{\zeta}_i B_i u_i)] \quad (37)$$

$$\mathcal{K}_{i+1} = \mathcal{P}_i\mathcal{C}_{i+1}^T(\mathcal{C}_{i+1}\mathcal{P}_i\mathcal{C}_{i+1}^T)^{-1} \quad (38)$$

$$\mathcal{P}_{i+1} = (I - \mathcal{K}_{i+1}\mathcal{C}_{i+1})\mathcal{P}_i. \quad (39)$$

To implement the preceding, we need the input matrix B_i and the covariance matrix $\mathcal{P}_0$. These can be approximated from N batch data according to

$$B_i = \left\{\begin{bmatrix} \mathcal{P}_{yu}\mathcal{P}_u^{-1} \\ \mathcal{P}_{qu}\mathcal{P}_u^{-1} \end{bmatrix}\right\}_i \approx \left\{\begin{bmatrix} Y \\ Q \end{bmatrix}U^T(UU^T)^{-1}\right\}_i \quad (40)$$

and

$$\mathcal{P}_0 = E\{\zeta_0 \zeta_0^T\} = \begin{bmatrix} \mathcal{P}_y & \mathcal{P}_{yq} \\ \mathcal{P}_{qy} & \mathcal{P}_q \end{bmatrix} - \begin{bmatrix} \mathcal{P}_{yu} \\ \mathcal{P}_{qu} \end{bmatrix} \mathcal{P}_u^{-1} [\, \mathcal{P}_{uy} \; \mathcal{P}_{uq}] \quad (41)$$

$$\approx \frac{1}{N} \left(\begin{bmatrix} Y \\ Q \end{bmatrix} [Y^T \; Q^T] - \begin{bmatrix} Y \\ Q \end{bmatrix} U^T (UU^T)^{-1} U [Y^T \; Q^T] \right). \quad (42)$$

As new measurement information becomes available, the predictions of the quality variables are recursively upgraded based on the correlation between the measured states and the quality variables that is captured in the sample covariance matrices. In this fashion, updated predictions of the final quality variables ($\hat{q}_i$) are obtainable at each time step on the basis of a *single* model. The quality predictions are simply the last n_q elements of the filter state estimate according to

$$\hat{q}_i = \underbrace{[0 \; I_{n_q}]}_{\mathcal{C}_q} \hat{\xi}_i + (B_q^+)_i \begin{bmatrix} u_i \\ \vdots \\ u_{t_f - 1} \end{bmatrix} \quad (43)$$

$$(B_q^+)_i = \left[(B_q)_i \cdots (B_q)_{t_f - 1} \right], \quad (44)$$

with the effects of the known future (from i to $t_f - 1$) input moves included. Thus, we have a statistically optimal general linear recursive predictor of final quality based on the measurements up to the current time *and* the known process inputs (given that N is large). The recursive solution is exactly equal to the batch least-squares solution once all the measurements become available.

Using the Kalman filter framework, the objective of recursive quality prediction has been achieved while maintaining the desired statistical interpretation. This approach avoids the need to create multiple models at several time instances, as well as the need to make assumptions about the nature of future measurements.

Comment. It is possible that B_i matrices obtained according to Eq. 40 show a causality (i.e., the future inputs have effect on the past inputs). This is clearly contrary to the physics. As an alternative strategy, one could first estimate $(B_y)_i$ and $(B_q)_i$ for $i = 1, \ldots, t_f$ from Y, Q, and U using least squares or any other method with causality enforced in the estimation. Once the effects of the deterministic inputs have been modeled, the residual data matrices ($\bar{Y}$ and $\bar{Q}$) are formed by subtracting off their effects (e.g., $\bar{\mathcal{Y}} = \mathcal{Y} - (\hat{B}_y^+)_0 \mathcal{U}$ and $\bar{q} = q - (\hat{B}_q^+)_0 \mathcal{U}$). The Kalman filter implementation is just as before with the initial state $[\bar{\mathcal{Y}}^T \; \bar{q}^T]_0^T$ and corresponding initial covariance given by

$$\mathcal{P}_0 = \begin{bmatrix} \frac{1}{N} \sum_{j=1}^{N} \bar{\mathcal{Y}}(j) \bar{\mathcal{Y}}^T(j) & \frac{1}{N} \sum_{j=1}^{N} \bar{\mathcal{Y}}(j) \bar{q}^T(j) \\ \frac{1}{N} \sum_{j=1}^{N} \bar{q}(j) \bar{\mathcal{Y}}^T(j) & \frac{1}{N} \sum_{j=1}^{N} \bar{q}(j) \bar{q}^T(j) \end{bmatrix} = \begin{bmatrix} \hat{\mathcal{P}}_{\bar{y}} & \hat{P}_{\bar{yq}} \\ \hat{\mathcal{P}}_{\bar{qy}} & \hat{\mathcal{P}}_{\bar{q}} \end{bmatrix} \quad (45)$$

based on the residual data (with input effects removed) for the past N batches.

Handling Colinearity and Reducing the Order of the Kalman Filter. At this point, we need to address a few problems associated with this approach. Since $\mathcal{Y}$ (especially) and q (possibly) are potentially of *very* large dimension, the covariances estimated from the data according to Eq. 42 may have large errors. The size of $\mathcal{Y}$ also results in a state dimension for the Kalman filter implementation that is potentially very large, thereby leading to large storage and computational requirements. In addition, colinearity in the data will probably cause the covariance matrices calculated through the approach in Eq. 42 to be ill-conditioned. One solution is to compute the sample covariance matrices according to:

$$\begin{bmatrix} \hat{\mathcal{P}}_y & \hat{\mathcal{P}}_{yu} & \hat{\mathcal{P}}_{yq} \\ \hat{\mathcal{P}}_{uy} & \hat{\mathcal{P}}_u & \hat{\mathcal{P}}_{uq} \\ \hat{\mathcal{P}}_{qy} & \hat{\mathcal{P}}_{qu} & \hat{\mathcal{P}}_q \end{bmatrix} = \begin{bmatrix} \frac{1}{N} \sum_{j=1}^{N} \mathcal{Y}(j) \mathcal{Y}^T(j) & \frac{1}{N} \sum_{j=1}^{N} \mathcal{Y}(j) \mathcal{U}^T(j) & P^T \left[\frac{1}{N} \sum_{j=1}^{N} \underline{\mathcal{Y}}(j) \underline{q}^T(j) \right] R \\ \frac{1}{N} \sum_{j=1}^{N} \mathcal{U}(j) \mathcal{Y}^T(j) & \frac{1}{N} \sum_{j=1}^{N} \mathcal{U}(j) \mathcal{U}^T(j) & \left[\frac{1}{N} \sum_{j=1}^{N} \underline{\mathcal{U}}(j) \underline{q}^T(j) \right] R \\ R^T \left[\frac{1}{N} \sum_{j=1}^{N} \underline{q}(j) \mathcal{Y}^T(j) \right] & R^T \left[\frac{1}{N} \sum_{j=1}^{N} \underline{q}(j) \underline{\mathcal{U}}^T(j) \right] P & \frac{1}{N} \sum_{j=1}^{N} q(j) q^T(j) \end{bmatrix}. \quad (46)$$

With this modification, formulating the Kalman filter based on the relevant covariance elements just as before leads to exactly the same quality prediction as $\hat{\Theta}^{BR}$ of Eq. 31 when all of the measurements become available at the end of the batch ($i = t_f$). Just as the previous filter formulation provided the recursive solution to the batch least-squares problem, this formulation provides the recursive solution to the batch biased regression problem. As more and more measurements become available during the course of the batch, the recursive solution more closely approaches the final batch solution.

Roughly speaking, the estimation works by using $\mathcal{P}_y$ to estimate the future measurements $y_{i+1}, \ldots, y_{t_f}$ from the currently available measurements $y_1, \ldots, y_i$. Subsequently, the final quality is predicted based on $y_1, \ldots, y_i, \hat{y}_{i+1}, \ldots, \hat{y}_{t_f}$ via biased regression. In concept, it is similar to the method proposed by Nomikos and MacGregor (1995a), who use the correlation given by the PLS model to estimate the future outputs. In the preceding method, however, estimates at the beginning of the batch are expected to be less vulnerable to measurement noises since the variations in $\mathcal{Y}$ other than those with correlation to q are modeled in the covariance matrix for $\mathcal{Y}$. On the other hand, $\mathcal{Y}$ is still of potentially

very large dimension, which leads to computational difficulties and conditioning problems.

A second option is to use the projected data for calculating *all* the covariances:

$$\begin{bmatrix} \hat{\mathcal{P}}_z & \hat{\mathcal{P}}_{zq} \\ \hat{\mathcal{P}}_{qz} & \hat{\mathcal{P}}_q \end{bmatrix} = \begin{bmatrix} P^T & 0 \\ 0 & R^T \end{bmatrix} \begin{bmatrix} \frac{1}{N}\sum_{j=1}^{N} \underline{\mathcal{Z}}(j)\underline{\mathcal{Z}}^T(j) & \frac{1}{N}\sum_{j=1}^{N} \underline{\mathcal{Z}}(j)\underline{q}^T(j) \\ \frac{1}{N}\sum_{j=1}^{N} \underline{q}(j)\underline{\mathcal{Z}}^T(j) & \frac{1}{N}\sum_{j=1}^{N} \underline{q}(j)\underline{q}^T(j) \end{bmatrix} \begin{bmatrix} P & 0 \\ 0 & R \end{bmatrix}. \tag{47}$$

The preceding may then be used as the basis of the recursive filter implementation. In this case, the dimension of the problem may be reduced by reformulating the Kalman filter with the $(n_{zr} + n_{qr}) \times 1$-dimensional *reduced* state $\underline{\zeta}_i$ (which is defined as $[\underline{\mathcal{Z}}^T \underline{q}^T]^T$ with $u_{i+j} = 0$ for $j \geq 0$) with the following state-space model:

$$\underline{\zeta}_{i+1} = \underline{\zeta}_i + \underline{B}_i u_i \tag{48}$$

$$y_i = \left[\underline{\mathcal{C}}_i \left[P^T \; 0_{n_{qr}}\right]\right] \underline{\zeta}_i + \epsilon_i = \underline{\mathcal{C}}_i \underline{\zeta}_i + \epsilon_i, \tag{49}$$

where $\underline{B}_i$ consists of the columns of $\mathcal{P}_{zu}\mathcal{P}_u^{-1}$ and $\mathcal{P}_{qu}\mathcal{P}_u^{-1}$ corresponding to the ith input. As will be discussed in the next paragraph, the addition of the measurement error term (ϵ_i) means that an additional term (R_ϵ^i) representing the covariance of ϵ_i must now be included in the Kalman gain matrix of the filter:

$$\underline{\mathcal{K}}_{i+1} = \underline{\mathcal{P}}_i \underline{\mathcal{C}}_{i+1}^T (\underline{\mathcal{C}}_{i+1} \underline{\mathcal{P}}_i \underline{\mathcal{C}}_{i+1}^T + R_\epsilon^i)^{-1}. \tag{50}$$

The reduction in filter dimension from this formulation constitutes a substantial computational savings while maintaining the well-conditioned properties associated with multivariate statistical model. The filter is initiated with $\hat{\underline{\zeta}}_0 = 0$ (the projected state with *no* input move effects) and the corresponding covariance matrix:

$$\underline{\mathcal{P}}_0 = \begin{bmatrix} \frac{1}{N}\sum_{j=1}^{N} \underline{\bar{\mathcal{Z}}}(j)\underline{\bar{\mathcal{Z}}}^T(j) & \frac{1}{N}\sum_{j=1}^{N} \underline{\bar{\mathcal{Z}}}(j)\underline{\bar{q}}^T(j) \\ \frac{1}{N}\sum_{j=1}^{N} \underline{\bar{q}}(j)\underline{\bar{\mathcal{Z}}}^T(j) & \frac{1}{N}\sum_{j=1}^{N} \underline{\bar{q}}(j)\underline{\bar{q}}^T(j) \end{bmatrix} = \begin{bmatrix} \hat{\mathcal{P}}_{\underline{z}} & \hat{\mathcal{P}}_{\underline{zq}} \\ \hat{\mathcal{P}}_{\underline{qz}} & \hat{\mathcal{P}}_{\underline{q}} \end{bmatrix}, \tag{51}$$

which is based on the *projected residual* data (with input effects removed as previously discussed) for the past N batches analogous to Eq. 45.

Note that a prediction error term ϵ has been included for the measurement equation in Eq. 49 for this formulation. This modification is necessary because the measurements are no longer the states as in the original formulation. Instead, the states are the *reduced* measurements (and quality), and some information is necessarily lost when using the multivariate statistical model to project down to the lower-dimensional space. The discarded information is not useful for prediction, so it is best treated as white noise. The covariance (R^ϵ) of the residual ϵ may be calculated based on the residuals of the available past data.

In the original formulation, the recursive filter solution will give exactly the batch solution once all of the data are available; however, this recursive multivariate statistical solution will not, in general, give the exact batch multivariate statistical solution because of the approximations made regarding the time correlation of the residual. On the other hand, provided that a sufficient number of reduced variables are maintained during the model-building step, the difference will hardly be significant.

Comment. When separate projections are used for the $\mathfrak{Y}$ and $\mathfrak{U}$ data

$$\begin{bmatrix} S & 0 \\ 0 & V \end{bmatrix} \begin{bmatrix} \mathfrak{Y} \\ \mathfrak{U} \end{bmatrix} \rightarrow \begin{bmatrix} \mathfrak{Y} \\ \mathfrak{U} \end{bmatrix} \tag{52}$$

the state-space model can be formulated as

$$\begin{bmatrix} \underline{\mathfrak{Y}} \\ \underline{q} \end{bmatrix}_{i+1} = \begin{bmatrix} \underline{\mathfrak{Y}} \\ \underline{q} \end{bmatrix}_i + \begin{bmatrix} B_{\underline{y}} \\ B_{\underline{q}} \end{bmatrix}_i u_i \tag{53}$$

$$y_i = \left[[0 \cdots I_{n_y} \cdots 0]\left[S^T \; 0_{n_q}\right]\right] \begin{bmatrix} \underline{\mathfrak{Y}} \\ \underline{q} \end{bmatrix}_i + \epsilon_i = \underline{\mathcal{C}}_i \begin{bmatrix} \underline{\mathfrak{Y}} \\ \underline{q} \end{bmatrix}_i + \epsilon_i \tag{54}$$

where

$$(B_{\underline{y}})_i = \mathcal{P}_{\underline{yu}} \mathcal{P}_{\underline{u}}^{-1} V([0 \cdots I \cdots 0])^T \tag{55}$$

$$(B_{\underline{q}})_i = \mathcal{P}_{\underline{qy}} \mathcal{P}_{\underline{u}}^{-1} V([0 \cdots I \cdots 0])^T \tag{56}$$

Again, a covariance term for ϵ_i must be added to the Kalman gain equation as before.

Assessing Model Validity. As previously stated, we are now in a position to use the Kalman filter theory to address the quality prediction problem. For instance, one may wish to monitor the normalized, squared filter prediction error $\epsilon_i^T \Sigma_i^{-1} \epsilon_i$ (where $\epsilon_i = y_i - \mathcal{C}_i z_i$ and Σ_i is its covariance matrix) in order to assess the performance of the filter. Under normal circumstances, the squared residual should exhibit χ^2 statistics for which normal bounds may be calculated using either theoretical distributions or the process data as discussed in Mehra and Peschon (1976). When the monitored residual of a new batch exceeds these bounds, it is an indication that a disturbance has occurred that represents an unmodeled change in the dynamics of the process. Thus, the

data-based model may no longer be valid and the resulting quality predictions become suspect.

In addition to the residual, the "size" of the filter state in a normal sense represents the "distance" from the region of normal operation and model validity since the state has been centered. This type of "drift" may be caused by a modeled disturbance that occurs at a magnitude not contained in the model-building data set. Although this type of drift from the normal operating region can be detected using the full data, an attractive approach to monitoring the size of the state involves the use of multivariate statistics as described in MacGregor et al. (1994) and Nomikos and MacGregor (1995b). According to these references, the basic idea is to project the data down to the reduced-dimensional space defined by the multivariate statistical model. In the reduced space, monitoring the "size" of the projection is performed using traditionally accepted statistical monitoring techniques based on the T^2 statistic (MacGregor et al., 1994; Nomikos and MacGregor, 1995b; Wise and Gallagher, 1997). As the sum of the normalized and squared variables in the reduced space, the T^2 statistic represents the magnitude of the variation with the confines of the statistical model. The monitoring procedure involves the comparison of the calculated T^2 value for the new batches to normal bounds for the statistic taken from theoretical distributions or the data themselves.

It is important to realize that we are essentially monitoring the validity of the model using these techniques. This is different from traditional process monitoring applications in which the goal is simply to detect any abnormal operation resulting from disturbances.

Data-Based Quality Control

The focus of this section is on developing an on-line data-based quality control framework that utilizes the control-relevant prediction models presented in the previous section. In this approach, the input moves necessary for obtaining the desired quality targets are calculated by solving a simple optimization problem based on the empirical prediction models. The discussion begins with a presentation of the chosen optimal control framework before moving on to the topics of directional input weighting, and input parameterization.

Formulation of the optimal control objective

The appropriate control objective for this type of system is to manipulate the input trajectory $\mathfrak{U}$ (which may serve as the set point for low-level controllers) in order to achieve the desired final product quality. An approach to realizing this goal is to solve for the remaining future input moves ($\mathfrak{U}^+ = u_i, \ldots, u_{t_f-1}$) at any time i ($0 \le i \le t_f - 1$) during the batch by solving the following general optimization problem:

$$\min_{\mathfrak{U}_i^+} \hat{q}_i^T \Lambda_q \hat{q}_i + (\mathfrak{U}_i^+)^T \Lambda_u \mathfrak{U}_i^+, \tag{57}$$

subject to constraints on the inputs and a quality prediction model such as

$$\hat{q}_i = \hat{\Theta}_i^{LS} \begin{bmatrix} \mathfrak{Y}_i \\ \mathfrak{U} \end{bmatrix}, \tag{58}$$

where

$$\mathfrak{U} = \begin{bmatrix} \mathfrak{U}_i \\ \mathfrak{U}_i^+ \end{bmatrix} \tag{59}$$

in the least-squares framework. In the Kalman filter framework, the prediction is given by

$$\hat{q}_i = R^T [0_{n_{sr}} \; I_{n_{qr}}] \hat{\underline{\zeta}}_{i|i} + \left[(B_{\underline{q}})_i \cdots (B_{\underline{q}})_{t_f-1} \right] \mathfrak{U}_i^+, \tag{60}$$

where $\hat{q}_{i|i}$ is available from the filter at time i. In general, Λ_q and Λ_u are weighting matrices, the elements of which are chosen according to the relative importance of the respective variables. Although linear model structures are the primary focus of this article, nonlinear regression models of the form in Eq. 8 may be required for strongly nonlinear systems. The computational demand of the optimization problem, which is quite small for linear models, will of course be increased considerably when using nonlinear regression models.

This type of controller is essentially a "shrinking horizon" model-predictive controller application similar to that described in Joseph and Hanratty (1993). While all of the remaining input moves are calculated at each point i, only the first of these is implemented, while the rest are recalculated at the next point. The shrinking-horizon terminology stems from the fact that the number of input moves remaining to be chosen decreases as time nears the end of the batch.

Directional input weighting for improved robustness

As discussed in previous sections, the empirical regression models are dependent on the data used in model building, and thus are only valid in a certain region defined by those data. If the controller specifies input moves that will take the process out of the region of model validity, the quality predictions used in the control calculation will no longer be reliable. This implies that one should try to obtain data that contain all the important disturbances and also all possible input move combinations (i.e., inputs that persistently excite the process). In practice, however, it may not always be feasible to collect such "rich" data.

The robustness problem that ensues may be partially addressed using the model assessment and monitoring techniques described in the second section. Recall that methods for monitoring the squared prediction error using χ^2 statistics and the drift from the region of normal operation using T^2 statistics were proposed as means of detecting operation in regions of poor model reliability. At any desired prediction point, these statistics may be calculated as functions of the currently available measurements *and* the manipulated inputs using the regression model. In addition, since the T^2 statistics of $\hat{\underline{z}}_{t_f|i}$ (or $\hat{z}_{t_f|i}$) are a function of the future input moves, they may be included in the controller optimization problem (Eq. 57) to ensure that the resulting input moves keep the process in the region of model validity. Thus, some measure of robustness to modeling errors is obtained.

The objective may be accomplished by incorporating the monitoring statistics into the optimization formulation in one of two ways. The first way is simply to introduce additional constraints into Eq. 57 that require the calculated monitoring

statistics to be less than their respective bounds, as determined from the model-building data set or theoretical distributions. Thus, the process is constrained to operate in the region of model validity. The second approach is to add additional terms to the objective function of Eq. 57 that represent the (weighted) value of the monitoring statistics as functions of the future input moves. For example:

$$\min_{\mathfrak{U}_i^+} \hat{q}_i^T \Lambda_q \hat{q}_i + (\mathfrak{U}_i^+)^T \Lambda_u \mathfrak{U}_i^+ + \Lambda_s T^2(\mathfrak{U}_i^+), \qquad (61)$$

where Λ_s is a user-chosen weight that defines the relative importance of the monitoring statistic in relation to the other controller objectives. While this formulation maintains penalties for digression from the model validity region, it has the advantage of reducing the chance of obtaining an infeasible solution that may occur frequently with the first method.

Input parameterization

At this point, the quality predictions and the control move calculations are based on the full $(n_u \cdot t_f) \times 1$ vector $\mathfrak{U}$. For many realistic systems, $\mathfrak{U}$ can be a very large vector. In this case, it is advantageous to parameterize the actual (not deviation) input vector $\mathfrak{U}$ with a much smaller vector α: $\mathfrak{U} = [u_1^T(\alpha), \ldots, u_{t_f-1}^T(\alpha)]^T$. Subsequently, one can express the model in terms of deviations in the input parameter vector of much smaller dimension (α' with the prime omitted hereafter). Based on an analysis of typical batch-process input profiles, a very useful parameterization approach can be achieved by discretizing the batch into several intervals and selecting parameters for each interval. The input parameterization that should be used for each interval depends on the length of the interval and sensitivity of the quality variables to high-frequency input changes during the interval. In the simplest case, each trajectory can be parameterized to be constant or have a constant slope during the interval (yielding one parameter). In practice, this is a very useful approach because many typical input profiles are piecewise linear. The other extreme is to parameterize with input values at all sample times present within the interval. The more parameters that are included, the "richer" the model identification data needs to be. Often, parameterization of very fine resolution is not needed; it will only make the identification more difficult.

Although input parameterization is extremely attractive from an identification standpoint, there are many benefits to be gained from a control perspective as well. Parameterization by the discretization approach advocated in the previous paragraph fits nicely in the optimal control framework used in this article. Because α is much smaller than $\mathfrak{U}$, the optimization in Eq. 61 becomes much simpler due to the smaller number of variables to be optimized:

$$\min_{\alpha_i^+} \hat{q}_i^T \Lambda_q \hat{q}_i + (\alpha_i^+)^T \Lambda_\alpha \alpha_i^+ + \Lambda_s T^2(\alpha_i^+), \qquad (62)$$

where α_i^+ represents the parameters related to the *future* inputs with respect to time i. Finally, the discretization approach allows multirate control to be implemented. Although measurements are taken at each instant, the control calculation need only be performed at the beginning of each interval to obtain the input parameters needed to take the process to the next interval.

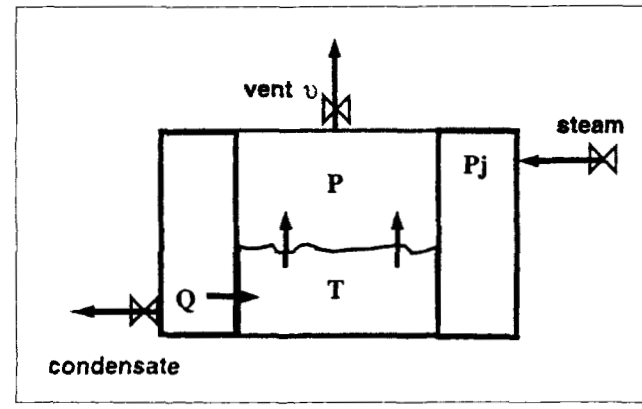

Figure 3. Typical nylon autoclave.

Application to a Nylon 6,6 Autoclave Reactor

Introduction to the process

The industrially relevant example considered in this article is the batch polymerization of nylon 6,6. More details regarding the model of the process can be found in Russell et al. (1998b). This step-growth polymerization involves the reversible reaction between hexamethylene diamine (HMD) and adipic acid monomers. Typically, the polymerization is performed in an autoclave reactor (see Figure 3) fitted with a valve for venting the water produced from the reaction. The reactor also has a steam jacket for supplying the heat needed for reaction and vaporization. Initially, an equimolar mixture of aqueous monomer salts (HMD and adipic acid) is fed to the autoclave from an evaporator. Once the mixture is charged, heat is supplied through the jacket to drive the polymerization reaction, with the vent being closed to prevent the loss of the volatile amine supplying monomer HMD. When most of the free HMD has reacted, the valve is opened to vent off the water, shifting the equilibrium of the polymerization toward the formation of higher molecular-weight polymer. Heating is then continued until the desired quality is achieved.

The relevant product quality in this process is characterized by the number average molecular weight (MW) and the concentration of amine end groups on the polymer molecules (NH2). Theoretically, the achievable molecular weight depends on r, the ratio of the monomers, and ϵ, the extent of the reversible polymerization reaction, as discussed in Schork et al. (1993). The addition of heat and the removal of water by venting drives the polymerization reaction toward the formation of higher molecular-weight polymer. The monomer HMD is also volatile, however, and its loss through vaporization reduces the ratio of monomers r and lowers the achievable molecular weight by a significant amount. From the standpoint of achieving target quality, it is critical to begin venting only after a sufficient amount of HMD has reacted (characterized by the extent of reaction ϵ).

Although measurements of MW and the amine end concentration NH_2 are made only at the end of the batch, on-line measurements of secondary process variables are typically available for monitoring and control purposes. These measurements include pressure readings for the reactor (P) and steam jacket (P_j), as well as temperature readings at the reactor center (T). In some cases, measurements of the vent flow rate (v) are also available. From a control standpoint, valves may be adjusted to manipulate the steam and reactor pressures. The steam pressure roughly determines the heat input (Q_{heat}) to the reactor, while the reactor pressure roughly determines the vaporization rate.

In addition to the lack of on-line measurements, the quality control problem is further complicated by the presence of frequent process disturbances. The disturbances to the process can be classified as either feed upsets or heat-transfer disturbances. Because one evaporator typically serves several autoclaves, scheduling problems result in variations in the feedwater content being a primary source of disturbance in this system. One of the major heat-transfer disturbances is reactor fouling, which occurs when polymer residue accumulates on the reactor walls between batches. This polymer film greatly reduces the efficiency of heat transfer to the reactor.

Traditional control approach

In the past, a common control strategy for nylon autoclaves has been to manipulate the vent and steam valves to follow nominal reactor and steam-jacket pressure profiles as shown in Figures 4 and 5 (Russell et al., 1998b). These nominal trajectories are predetermined from historical data of acceptable batches; they are normally not updated to account for changing process conditions. The pressure profile control approach is attractive because of the ready availability of pressure measurements as well as the simplicity of constructing effective PID control loops to track the pressure profiles, using the valve positions as inputs. This method will serve as the base case for comparing the performance of the data-based quality control approach.

Basis of the prediction and control problem

In this section, we set the stage for the application of the proposed data-based prediction and control approach to the nylon polymerization system. For this application, the on-line measurements include center temperature and vent-rate readings taken at 15-s intervals (typically 700 to 800 intervals per batch). The manipulated inputs are the set point trajectories (P_j and P) for the PID pressure controllers given as the base case PID controllers discussed in the previous paragraph. It is assumed that the performance of the tracking controllers is such that the pressures themselves may be taken as the inputs. The process disturbances to be rejected are fluctuations in the feedwater content ($\pm 10\%$) and the level of reactor fouling (0 to 10%).

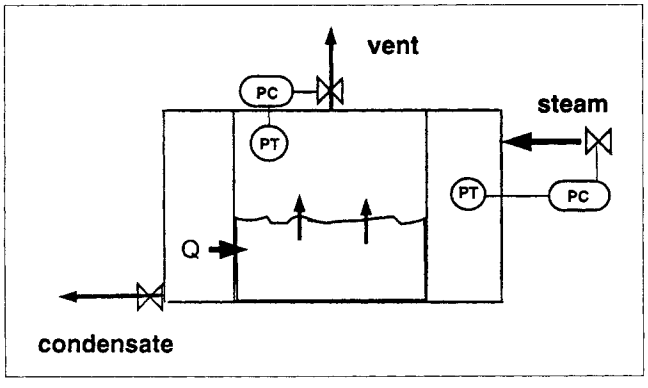

Figure 4. Reactor and steam-jacket pressure-control configuration.

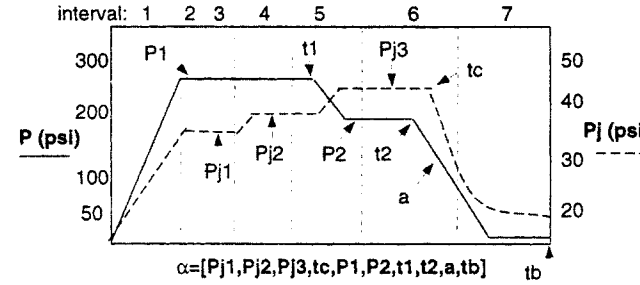

Figure 5. Application of discretization and parameterization.

Because of the potentially very large input dimension, an application of the parameterization strategy of the third section is used for this example. An analysis of the typical pressure profiles for the autoclave indicates that parameterization using approximately piecewise-linear profiles, as shown in Figure 5, should be sufficient. Based on process knowledge, it is apparent that the key operating region occurs around the time the vent valve is opened to begin venting the excess water. Therefore, the batch is discretized using a total of seven intervals, with a higher frequency of intervals just before and just after venting normally starts (about 35 min into the batch). On the other hand, late input moves have little effect on the final quality, meaning that a low interval frequency is sufficient toward the end of the batch. A total of 10 input parameters (α) are chosen as the various pressure levels, rates of change, and time instances necessary to construct typical piecewise-linear pressure profiles, some of which are shown in Figure 5. The steam-pressure trajectory parameters include the pressure levels ($Pj1$, $Pj2$, and $Pj3$) and the steam cutoff point (tc), after which no more steam is added. The reactor pressure trajectory parameters include the pressure levels ($P1$ and $P2$), the pressure-decrease points ($t1$ and $t2$), and the slope of final pressure decrease (a), as well as the final batch time (tb). Typical constraints are added for each of these parameters to reflect normal operation and to ensure that the time-related parameters ($t1$, $t2$, tc, and tb) each remain inside their respective control interval.

Developing quality prediction models

The first step in constructing a quality prediction model for this system is to gather the appropriate control-relevant, model-building data set. A total of 69 batch simulation runs containing the typical feed and heat-transfer disturbances as well as input moves were selected for this purpose. The disturbance cases were generated by randomly varying the ini-

tial water content and heat-transfer coefficient between typical limits for each simulation run. Following the discussion of the second section, the first approach we will consider is the multiple-model approach, where separate regression models are developed for the beginning of each discretization interval. For each interval, models of the form in Eq. 9 are built using the measurement data only up to the beginning of the interval, as well as the input and quality measurement data from the model-building data set. Because of the large number (~ 1,500 per batch) and high degree of correlation of the variables in the data set, the regression models were formulated using partial least squares (PLS), which reduces the number of variables down to about five to ten latent variables for each model. Control limits for the T^2 and prediction error monitoring statistics discussed in the second section were also obtained from the data for use in the assessment of model validity according to the procedures used in references previously discussed (MacGregor et al., 1994; Nomikos and MacGregor, 1995b; Mehra and Peschon, 1976). For comparison, the recursive Kalman filter predictor approach of Eqs. 53–54 in the second section has also been implemented on the basis of a PLS model with eight latent variables. Finally, the validity of all of the data-based models was confirmed using an independent model-testing data set.

Control strategy formulation

To perform on-line quality control, a shrinking-horizon MPC implementation of the type discussed in the third section was implemented on the nylon system. The basis of this implementation is an optimization problem such as in Eq. 62:

$$\min_{\alpha_i^+} \quad \hat{q}_i^T \Lambda_q \hat{q}_i + (\alpha_i^+)^T \Lambda_\alpha \alpha_i^+ + \Lambda_s T^2(\alpha_i^+), \qquad (63)$$

to be solved on-line at the beginning of each interval for the remaining (future) input parameters in α_i^+. Bounds on the various input parameters were also included in the optimization problem to reflect *a priori* knowledge of the limits of normal operation. For this example, $\Lambda_q = I$ was chosen because of the equal importance placed on the two normalized quality parameters (MW and NH_2). Because of the relatively low importance of the magnitude of the input moves within the stated bounds, $\Lambda_u = 0$ was chosen. According to the directional input weighting discussion of the third section, the T^2 statistic of the PLS model was included in the optimization problem to provide robustness to disturbances causing model errors. The relative weight (Λ_s) on the T^2 statistic was tuned by simulation experiments performed on the independent model-testing data set. Finally, the quality predictions used in the control calculations may be supplied by either the recursive or multiple-model data-based prediction approaches outlined in the previous paragraph.

Simulation results

To investigate the performance of the proposed data-based quality prediction and control approach for this example, the process was simulated 17 times with randomly varying feedwater content and reactor-fouling disturbances. When not controlled, 12 of these batches show unusual variation when compared to the statistical monitoring limits (prediction error and T^2) of the model-building data set, meaning that the data-based models will have to perform some extrapolation beyond their range of strict validity. The first objective of the simulation study is to assess and compare the predictive capabilities of the recursive and multiple-model methods. The second objective is to assess the performance of the quality-control strategy and investigate the effectiveness of the statistically based directional input weighting technique. Finally, the ability of the data-based models to detect abnormally operating batches outside the range of model validity using prediction error and T^2 statistics will be examined.

Prediction. In order to compare the multiple-model and recursive prediction methods and assess their predictive capability, the quality-prediction results for the 17 simulated test batches are shown in Figures 6–8 and Table 1 in terms of the deviation from the nominal value. For illustrative purposes, only the final model at interval 7 of the multiple-model approach will be compared to the recursive approach based on all the data up to the seventh interval; however, the re-

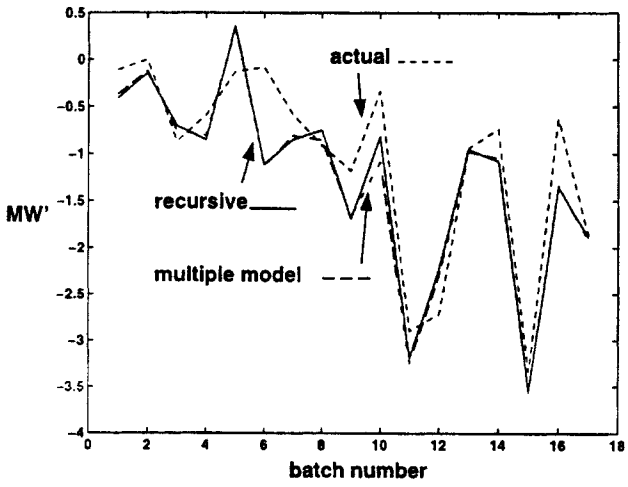

Figure 6. Number average molecular weight predictions for the test batches.

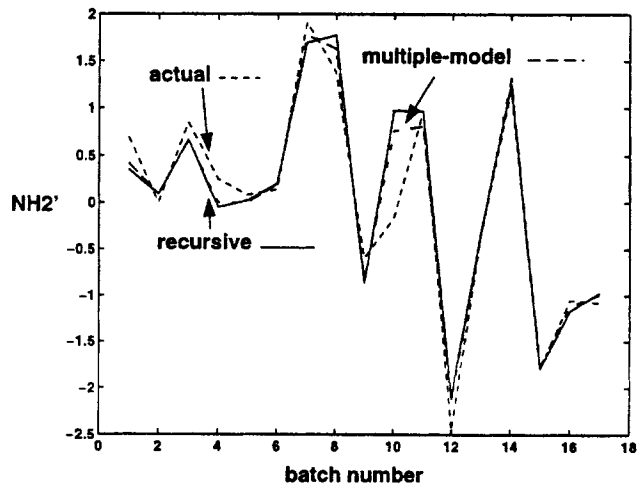

Figure 7. Amine end predictions for the test batches.

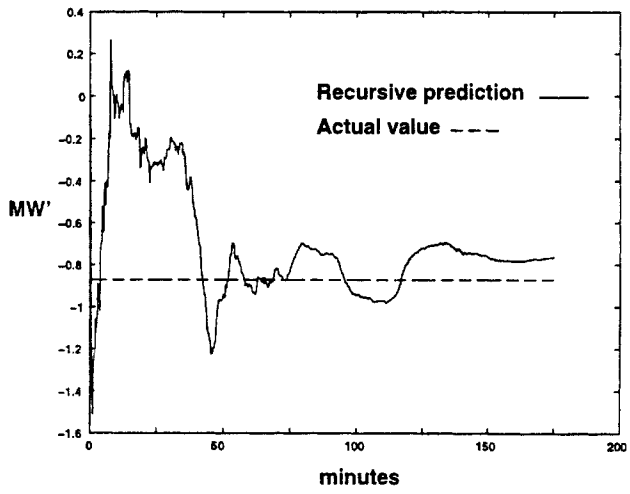

Figure 8. Time evolution of the recursive predictions within a batch.

sults using the models at the other intervals are very similar. Figures 6 and 7 demonstrate the similarity in predictive performance of the two approaches. These results are further verified by the sum of the squared quality-prediction errors for the 17 batches contained in Table 1. Figure 8 gives an indication of how the recursive filter predictions evolve throughout the batch as more measurements become available. As can be inferred from the figures, the prediction capability of the empirical models is quite satisfactory. Since both approaches advocated in this article show similar predictive capability, either may be used in the subsequent quality-control analysis.

Control and Model Assessment. The average and mean-square control error results for the data-based control strategy shown in Eq. 63 are compared to the traditional pressure control approach in Table 2 and Figures 9 and 10. These results show that the proposed approach offers much tighter quality control when compared to the desired limits, even though the data-based model is forced to extrapolate in many cases. The ability of the proposed control approach to counter process upsets by manipulating the input profiles is demonstrated in Figure 10. Note the relatively higher steam pressure that the data-based controller uses to drive the reaction in the presence of the excess water caused by the feed disturbance and the corresponding higher reactor pressure levels to prevent premature vaporization of the HMD monomer.

Recall that the objective function in Eq. 63 is the weighted sum of the predicted final quality deviations and directionally weighted inputs based on the predicted T^2 statistic. In Table 3, the performance of the control algorithm is compared for the following two cases: (1) weighting on the predicted quality deviations only ($\lambda_q = I$, $\lambda_u = 0$, $\lambda_s = 0$), and (2) weighting on the predicted quality deviations and the T^2 statistic ($\lambda_q = I$, $\lambda_u = 0$, $\lambda_s = 5/T^2_{\text{lim}}$, where T^2_{lim} is the desired bound on the T^2 statistic obtained from the model-building data set). As can been in Table 3, the addition of the directional input weighting improves the performance of the controller as well as its robustness in terms of the number of batches producing abnormal operating "alarms." For the purpose of this discussion, an "alarm" simply means that the respective monitoring statistic (prediction error or T^2) indicated abnormal operation for that batch relative to the 95% control limits of the model-building data set. An example of a statistical monitoring alarm-detecting operation in a region of poor model reliability is demonstrated in Figure 11, compared to a normal batch in Figure 12. The unusual variation caused by an abnormally large disturbance is easily detected in Figure 11 using the statistical tests by noting that the calculated T^2 statistic for the current batch at interval 3 exceeds the 95% confidence bounds based on the model-building data set. In addition, the prediction error statistic for the current batch exceeds its limits for intervals 4–6.

From these simulation results, one can conclude that the common pressure-control approach is not very robust to common process disturbances. On the other hand, the data-based appraoch presented in this article offers an attractive alternative due to its demonstrated effectiveness, robustness to disturbances, and low on-line computational demand.

Table 1. Sum of Squared Quality Prediction Error Results for the Test Batches

Method	SSPE MW	SSPE NH_2
Multiple-model @ 7	2.9	3.1
Recursive	3.2	1.3

Conclusions

In this article, a general data-based approach to obtaining on-line predictions of final product quality has been developed. This general formulation allows prediction of product quality under disturbance and control effects by the selection of an appropriate model-building data set. A novel recursive quality prediction formulation has also been developed on the basis of Kalman filter theory. This approach uses a single regression model to make statistically optimal recursive predictions of the final product quality at any point during the batch using only the on-line measurements available up to that point. In addition, multivariate statistical regression models have been incorporated into the proposed framework. These model structures give certain advantages in terms of dimension reduction and numerical conditioning while allowing connections to be made to the majority of the available literature on data-based batch process monitoring. Finally, concepts from Kalman filter theory and statistical batch-process monitoring have been included to detect when the regression models may no longer be valid.

Table 2. Quality Control Results for the Simulations

Method	Target MW	Target NH_2	Avg. MW	Avg. NH_2	MW MSE	NH_2 MSE
Press. Con.	13,447	49.7	12,997	45.2	7.2×10^6	953
Data-based	13,447	49.7	13,452	50.4	1.4×10^5	42.4

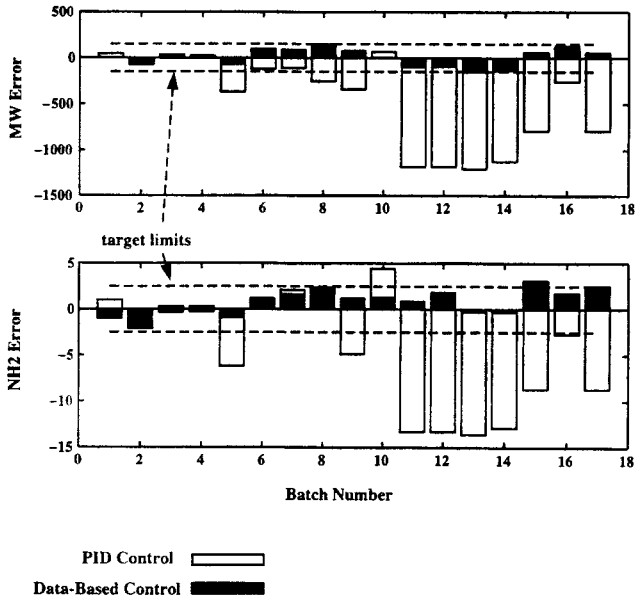

Figure 9. Quality control results for simulations.

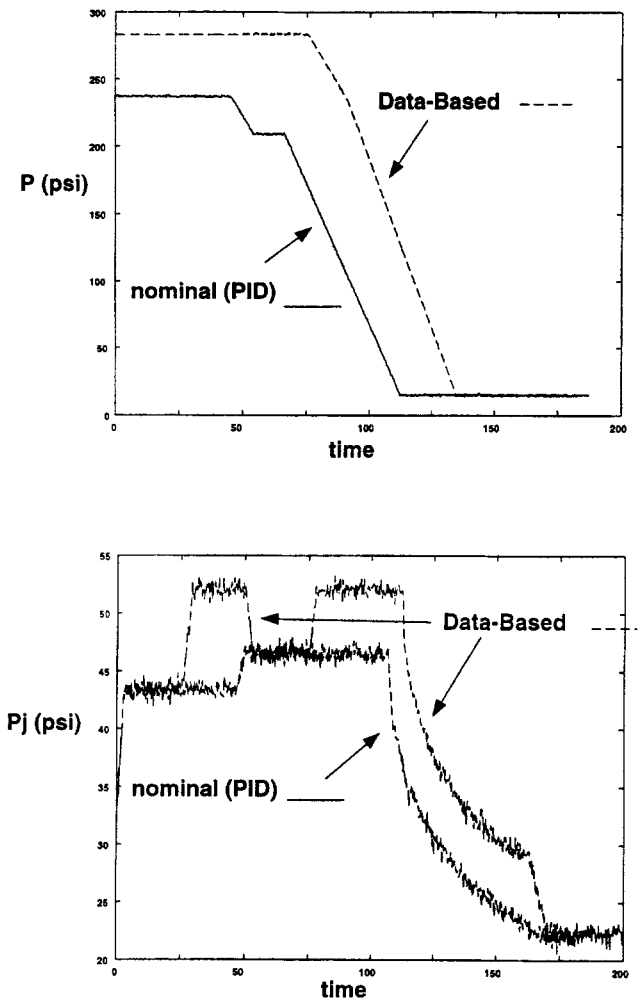

Figure 10. Example pressure profiles for a significant feed disturbance.

Table 3. Simulation Analysis of Statistically Based Directional Input Weighting

Case	Avg. MW Error	Avg. NH_2 Error	MW MSE	NH_2 MSE	T^2 Alarms	Pred. Err. Alarms
1	−107	6.3	1.6×10^6	119	13	15
2	5	0.7	1.4×10^5	42.4	4	7

The data-based quality-prediction approach provides the framework necessary for addressing the problem of batch quality control when on-line measurements of product quality are not available and the process experiences frequent

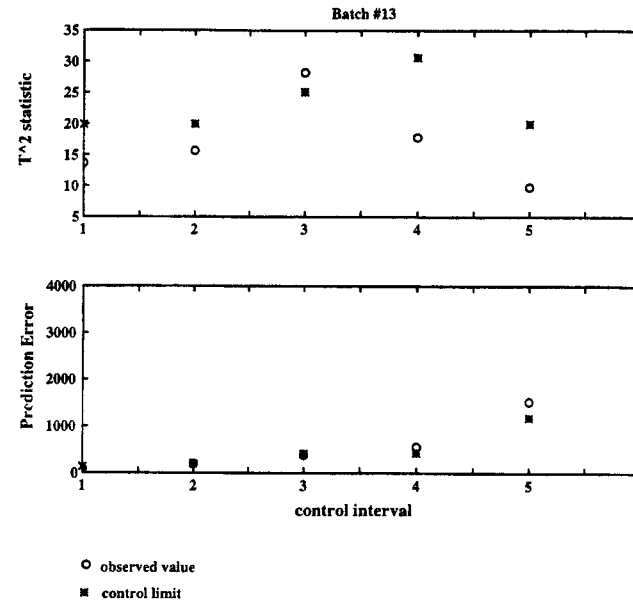

Figure 11. Example of the detection of an abnormally operating batch.

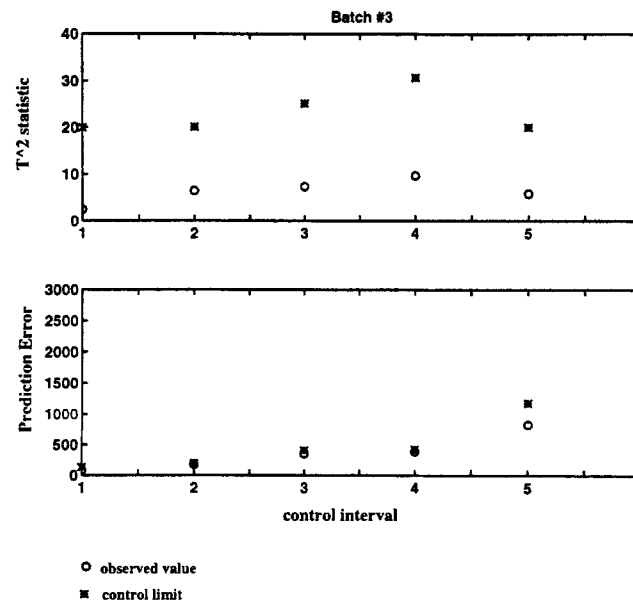

Figure 12. Example of a normally operating batch.

batch-to-batch variations in operating conditions. In this article, the proposed data-based prediction methods have been incorporated into an optimal inferential control strategy for product quality. The proposed control approach can be viewed as shrinking-horizon model-predictive control based on empirical models. Statistical monitoring concepts have been combined into the controller calculations via directional input weighting to provide robustness to modeling errors. The proposed control approach has the advantage of not relying on any fundamental process model and has a relatively mild on-line computational demand. Finally, the effectiveness of the proposed prediction and control methods was illustrated by an application to an industrially relevant simulated polymerization example.

Acknowledgment

The authors gratefully acknowledge the support of DuPont and the National Science Foundation (Grant CTS-9357827).

Literature Cited

Budman, H., T. Holcomb, and M. Morari, "PLS-Based Robust Inferential Control for a Packed-Bed Reactor," *ACC Conf. Proc.*, Boston, MA, Vol. 1, p. 256 (1991).

Budman, H., C. Webb, T. Holcomb, and M. Morari, "Robust Inferential Control for a Packed-Bed Reactor," *Ind. Eng. Chem. Res.*, **31**, 1665 (1992).

Cheng, G., and T. McAvoy, "Process Control Utilizing Data Based Multivariate Statistical Models," *Can. J. Chem. Eng.*, **74**, 1010 (1996).

Cheng, G., and T. McAvoy, "Multiblock Predictive Monitoring of Continuous Processes," *IFAC-ADCHEM Preprints*, p. 73 (1997).

Dong, D., and T. J. McAvoy, "Multi-Stage Batch Process Monitoring," *ACC Conf. Proc.*, Seattle, WA, p. 1857 (1995).

Dong, D., T. McAvoy, and E. Zafiriou, "Batch-to-Batch Optimization Using Neural Network Models," *Ind. Eng. Chem. Res.*, **35**, 2269 (1996).

Geladi, P., and B. Kowalski, "Partial Least-Squares Regression: A Tutorial," *Anal. Chim. Acta*, **185**, 1 (1986).

Gelb, A., ed., *Applied Optimal Estimation*, MIT Press, Cambridge, MA (1974).

Hoskuldsson, A., "PLS Regression Methods," *J. Chemometrics*, **2**, 211 (1988).

Jazwinski, A. H., *Stochastic Processing and Filtering Theory*, Academic Press, San Diego (1970).

Joseph, B., and P. Hanratty, "Predictive Control of Quality in a Batch Manufacturing Process Using Artificial Neural Network Models," *Ind. Eng. Chem. Res.*, **32**, 1951 (1993).

Kaspar, M. H., and W. H. Ray, "Chemometric Methods for Process Monitoring and High-Performance Controller Design," *AIChE J.*, **38**, 1593 (1992).

Kaspar, M. H., and W. H. Ray, "Dynamic PLS Modelling for Process Control," *Chem. Eng. Sci.*, **48**, 3447 (1993a).

Kaspar, M. H., and W. H. Ray, "Partial Least Squares Modelling as Successive Singular Value Decomposition," *Comput. Chem. Eng.*, **17**, 985 (1993b).

Kourti, T., and J. F. MacGregor, "Process Analysis, Monitoring, and Diagnosis, Using Multivariate Projection Methods," *Chemometrics Intelligent Lab. Syst.*, **28**, 3 (1995).

Kresta, J., J. MacGregor, and T. Marlin, "Multivariate Statistical Monitoring of Process Operating Performance," *Can. J. Chem. Eng.*, **69**, 35 (1991).

Ku, W., R. Storer, and C. Georgakis, "Disturbance Detection and Isolation by Dynamic Principal Components," *Chemometrics Intelligent Lab. Syst.*, **30**, 179 (1995).

Larimore, W. E., "Optimal Reduced Rank Modeling, Prediction, Monitoring, and Control Using Canonical Variate Analysis," *IFAC-ADCHEM Preprints*, p. 61 (1997).

Lee, J. H., and A. K. Datta, "Nonlinear Inferential Control of a Batch Pulp Digester," *AIChE J.*, **40**, 50 (1994).

MacGregor, J. F., T. Marlin, J. Kresta, and B. Skagerberg, "Multivariate Statistical Methods in Process Analysis and Control," *AIChE Symp. Proc. Int. Conf. on Chemical Process Control*, Y. Arkun and W. H. Ray, eds., p. 79 (1991).

MacGregor, J. F., C. Jaeckle, C. Kiparissides, and M. Koutoudi, "Process Monitoring and Diagnosis by Multiblock PLS Methods," *AIChE J.*, **40**, 826 (1994).

MacGregor, J., T. Kourti, and P. Nomikos, "Analysis, Monitoring, and Fault Diagnosis of Industrial Processes Using Multivariate Statistical Projection Methods," *Proc. IFAC World Cong.*, **M7a-05**(1), San Francisco (1996).

Mehra, R., and J. Peschon, "An Innovations Approach to Fault Detection and Diagnosis in Dynamic Systems," *Automatica*, **7**, 637 (1976).

Nomikos, P., and J. F. MacGregor, "Monitoring Batch Processes Using Multiway Principal Component Analysis," *AIChE J.*, **40**, 1361 (1994).

Nomikos, P., and J. MacGregor, "Multi-way Partial Least Squares in Monitoring Batch Processes," *Chemometrics Intelligent Lab. Syst.*, **30**, 97 (1995a).

Nomikos, P., and J. MacGregor, "Multivariate SPC Charts for Monitoring Batch Processes," *Technometrics*, **37**, 41 (1995b).

Ott, E. R., and E. G. Schilling, *Process Quality Control*, 2nd ed., McGraw-Hill, New York (1990).

Piovoso, M., K. Kosanovich, and R. Pearson, "Monitoring Process Performance in Real-Time," *ACC Conf. Proc.*, Chicago, p. 2359 (1992).

Piovoso, M., and K. Kosanovich, "Applications of Multivariable Statistical Methods to Process Monitoring and Controller Design," *Int. J. Control*, **59**, 743 (1994).

Qin, S., and T. McAvoy, "Nonlinear PLS Using Neural Networks," *Comput. Chem. Eng.*, **16**, 379 (1992).

Qin, S., and T. McAvoy, "Nonlinear FIR Modeling Via a Neural Net PLS Approach," *Comput. Chem. Eng.*, **20**, 147 (1996).

Robertson, D., J. Lee, and J. Rawlings, "A Moving Horizon-Based Approach for Least Squares Estimation," *AIChE J.*, **42**, 2209 (1996).

Russell, S., D. Robertson, J. Lee, and B. Oggunnaike, "Model-Based Monitoring of Product Quality for Batch and Semi-Batch Processes," *J. Proc. Cont.*, in press (1998a).

Russell, S. A., D. G. Robertson, J. H. Lee, and B. A. Ogunnaike, "Control of Product Quality for Batch Nylon 6,6 Autoclaves," *Chem. Eng. Sci.*, in press (1998b).

Schork, F. J., P. B. Deshpande, and K. W. Leffew, *Control of Polymerization Reactors*, Dekker, New York (1993).

Schuler, H., and Z. Suzhen, "Real-Time Estimation of the Chain Length Distribution in a Polymerization Reactor," *Chem. Eng. Sci.*, **40**, 1891 (1985).

Wise, B., and N. Gallagher, "PLS Toolbox," *MATLAB™ Toolbox Manual*, Eigenvector Technologies, Manson, WA (1997).

Wold, S., A. Ruhe, H. Wold, and W. J. Dunn III, "The Collinearity Problem in Linear Regression: The Partial Least Squares (PLS) Approach to Generalized Inverses," *SIAM J. Sci. Stat. Comput.*, **5**, 735 (1984).

Yabuki, Y., and J. MacGregor, "Product Quality Control in Semi-Batch Reactors Using Mid-Course Corrections Policies," *IFAC-ADCHEM Preprints*, p. 189 (1997).

Ye, N., T. McAvoy, K. Kosanovich, and M. Piovoso, "Plant-Wide Control Using an Inferential Approach," *ACC Conf. Proc.*, p. 1900 (1993).

Manuscript received Feb. 25, 1998, and revision received Sept. 10, 1998.