Troubleshooting of an Industrial Batch Process Using Multivariate Methods

Salvador García-Muñoz, Theodora Kourti, and John F. MacGregor*

Department of Chemical Engineering, McMaster University, Hamilton, Ontario, Canada L8S4L7

Arthur G. Mateos and Gerald Murphy

FMC Corporation, Chemical Products Group, Box 8, Princeton, New Jersey 08543

Multivariate statistical methods are used to analyze data from an industrial batch drying process. The objective of the study was to uncover possible reasons for major problems occurring in the quality of the product produced in the process. Partial least-squares (PLS) methods were able to isolate which group of variables in the chemistry, in the timing of the various stages of the batch, and in the shape of the time-varying trajectories of the process variables were related to a poor-quality product. The industrial study illustrates the approach and the power of these multivariate methods for troubleshooting problems occurring in complex batch processes. Several new variations in the multivariate PLS methodology for the analysis of batch data are also implemented. In particular, an application utilizing a novel approach to the time warping of the trajectories for batches, and the subsequent use of the time-warping information, is presented. The use of the time history of the PLS weights of the process variable trajectories to diagnose problems in the dynamic operation of the batches is also clearly illustrated, as is the use of contribution plots for finding features which distinguished between the operating histories of good and bad batches.

1. Introduction

Batch processes play an important role in the pharmaceutical, semiconductor, polymer, and specialty chemical industries. Data collected over the duration of the batch have a time dimension for each process variable. As a result, a data set from a batch process can be arranged in a three-dimensional array (Figure 1) consisting of data from a number of batches (I), with J process variables measured at K time intervals over the course of each batch.

Nomikos and MacGregor presented methods for applying multivariate statistical process control (MSPC)[1−3] by using multiway principal component analysis (PCA)[5] to an "unfolded" matrix of the batch data (Figure 1). This strategy models the deviations of each batch from the mean trajectory because it mean-centers the unfolded matrix columnwise and scales each column to have a variance of one. These multiway statistical models capture the correlation structure among all of the measured variables over the entire time history of the batch and compress this information down into a low-dimensional latent variable space, in which it was easy to compare and monitor batches. Nomikos and MacGregor[4] extended the methods using multiway partial least squares (PLS)[5] to incorporate the final product quality data (Y) collected at the end of each batch. Kourti et al.[6] further extended the method to incorporate data on the initial conditions and other discrete variables, such as operators and information from upstream processes, by using multiblock, multiway PCA and PLS (Figure 2).

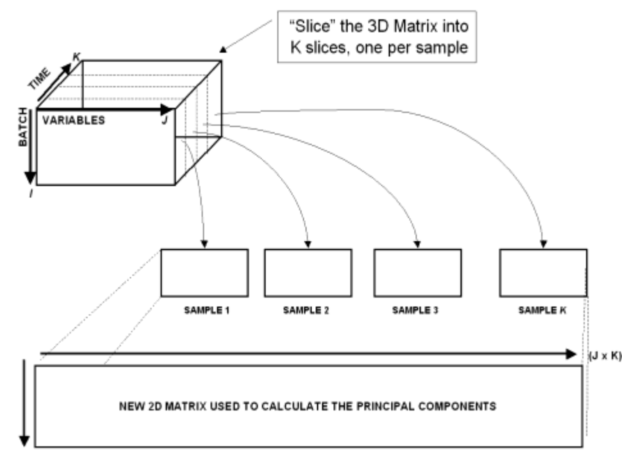

Figure 1. Unfolding of the three-way batch data set.

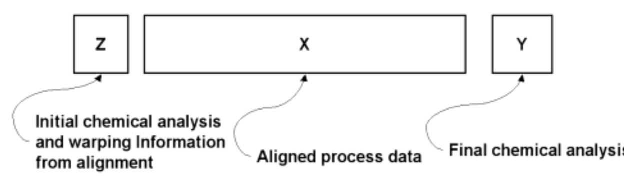

Figure 2. Multiblock multiway PLS matrices.

The methods proposed by Nomikos and MacGregor have been applied widely in industry, with some applications reported in the literature,[7−11] but most are still unpublished.

Other variations of these multivariate methods for monitoring batch systems have also been reported.[12,13]

The paper is organized as follows: Section 2 describes the process, its operation, and its variables and how the product is classified into on-spec and off-spec products

* To whom correspondence should be addressed. Tel.: 905 5259140 ext. 24951. Fax: +1 (905) 521 13 50. E-mail: macgreg@mcmaster.ca.

10.1021/ie0300023 CCC: $25.00
Published on Web 06/20/2003

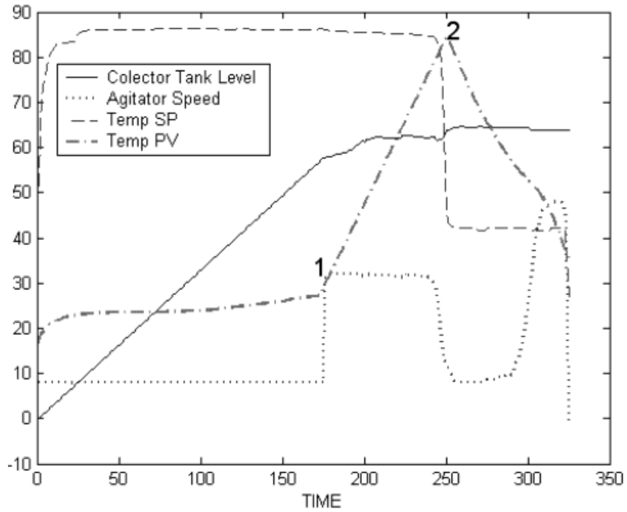

Figure 3. Critical variables and two critical steps in operation (level in percent fill, temperature in degrees Celsius, speed in rpm).

using PCA on the product quality data. Section 3 describes how the batch trajectories are synchronized and presents a new approach to utilizing the time-warping information from the synchronization. Section 4 describes the PLS modeling studies performed on the data and the use of these models to uncover the major operating problems with the process that are related to the production of a poor-quality product. Some conclusions are then given in section 5.

2. The Drying Process

The unit is used to evaporate and collect the solvent contained in the initial charge (wet cake) and to dry the product to a target residual solvent level. During the drying, important chemical structural changes can occur that can lead to unacceptable product quality. With the aid of Figure 3, the operation is described as follows:

1. The batch is charged with a mass of wet cake that varies from batch to batch. The weight of the wet cake fed to the dryer is measured for each batch, but the solvent content is unknown.

2. At the beginning of the batch, the agitator is running at low speed ($\sim$8 rpm) and the heating jacket is already running with a hot medium. As a result, the temperature in the dryer starts increasing slowly.

3. At a certain point determined by the control system (marked as “1” in Figure 3), the agitator is switched to high speed ($\sim$30 rpm) and the temperature increases rapidly until it reaches its peak (point “2” in Figure 3). The agitator is triggered down to slow speed just before point “2” is reached.

4. After the temperature peak (point “2”), the product in the batch is cooled, and then toward the end of the batch, the agitator is turned to high speed for some time.

As the batch evolves, the solvent that is being extracted from the wet cake is collected in a tank that is emptied after each batch. The time at which the agitator triggers from low to high speed and vice versa is different for each batch; also the maximum temperature that the batch reaches is not the same for all batches because the operators adjust this set point (peak temperature) from time to time to correct the quality of the product in a batch-to-batch manual feedback.

2.1. The Data. There are three sets of variables measured for each batch: (a) an initial chemical analysis on the wet cake done before each batch and the weight of the wet cake fed to the dryer (11 chemical variables and one weight), (b) 10 process variable trajectories as they evolve throughout the batch, and (c) 11 product quality variables measured at the end of the batch.

The data on the initial chemical analysis and on the weight of the cake will be referred to as the initial condition matrix (**Z**), the process variable trajectories will be referred to as the process matrix (**X**), and the final chemical properties will be referred to as the quality matrix (**Y**). A more detailed description as well as the names used for each measured variable is presented in Table 1. The whole data set consists of 71 batches.

2.2. Multivariate Classification of Quality. The quality of a product has to be analyzed in a multivariate way because “quality is a multivariate property, requiring the correct combination of all the measured characteristics”.[14] Therefore, to uncover the natural multivariate product classification within the data, we first build a PCA model on the final product properties (**Y**).

After removal of 12 obvious outliers, a final PCA model for **Y** with two components captures 70.0% of the data (51.2% captured by the first component and 18.8% by the second component). The t_1-t_2 score plot (Figure 4) shows a natural clustering of what the company had classified as on-spec and off-spec products; the dashed line ellipse represents the approximate 99% confidence interval in this model and contains the 59 used observations. The dashed line is drawn to show the separation between good- and poor-quality products, and it clearly shows the separation between good- and bad-quality products much more clearly than looking at the 11 **Y** variables separately. The model can be easily used to discriminate the quality of the final product by its position in the t_1-t_2 plane. Figure 5 shows the square prediction error (residuals of the PCA model) for each batch, together with their 95% and 99% confidence limits.[3] Aside from batch 41, the PCA model appears to explain the quality variation from the batches.

3. Alignment of the Batch Data

One of the issues still under investigation in the application of multivariate methods for batch process analysis is how to handle a set of batches with different durations; this is referred as the “alignment” or “synchronization” problem. To “align” or “synchronize” a data set of batch trajectories means to perform a certain transformation on each trajectory in such way that at the end of the “alignment” all of the batch trajectories will line up and “evolve” similarly and have the same number of samples.

3.1. Alignment Technique. In this industrial drying process, the amount of wet cake, the initial concentration of the solvent, and some operating conditions for each of the batches are different. As a result, the duration of each batch and the duration of different stages within each batch are different; thus, there is a need for performing an alignment on the trajectories in order to make them comparable. Nomikos and MacGregor[3] proposed the usage of an indicator variable in order to resample the trajectories and perform the alignment (e.g., sample at 1% increments on conversion). Kassidas et al.[15] proposed the use of dynamic time warping, a technique used in speech recognition, to align the batch data when an indicator variable is not obvious.

Table 1. Variables Measured Per Batch

matrix	variable name	description
Z	WgtCake	total weight of the wet cake fed initially to the dryer
Z	Z1, Z2, ..., Z10	nine organic group concentrations and pH
X	CTANKLVL	level of the collector tank; always starts at zero (empty)
X	DIFF-PRESS	differential pressure in the dryer
X	X1	pressure in the dryer
X	X2	power to the agitator
X	X3	torque resistance for the agitator
X	AGITSPEED	agitator speed
X	JTEMPsp	set point for the jacket heating medium
X	JTEMP	temperature of the jacket heating medium
X	DTEMPsp	set point for the temperature inside the dryer
X	DTEMP	temperature inside the dryer
Y	Y1, Y2, ..., Y10	seven organic group concentrations and three physical properties
Y	SOLVENT-CONC	weight percentage of solvent in the final product

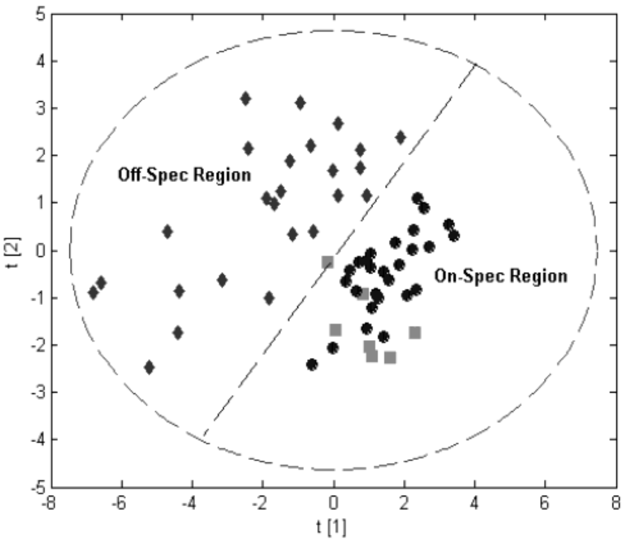

Figure 4. PCA score plot for product quality data (**Y**): off-spec (◆), on-spec (●), on-spec but high residual solvent (■).

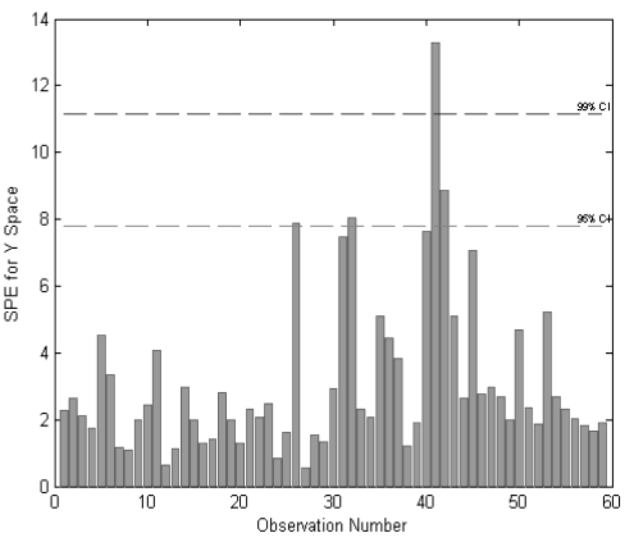

Figure 5. SPE for the **Y** matrix multivariate classification model.

In these two suggested techniques, the authors emphasize the importance of not discarding the information that comes out of an alignment (e.g., the time required to complete any given stage). This information will be called "warping information" because it is a new piece of data for each batch, and it is not available until the alignment is performed.

There are many ways to use an indicator variable to align batch trajectories. If there exists a monotonically

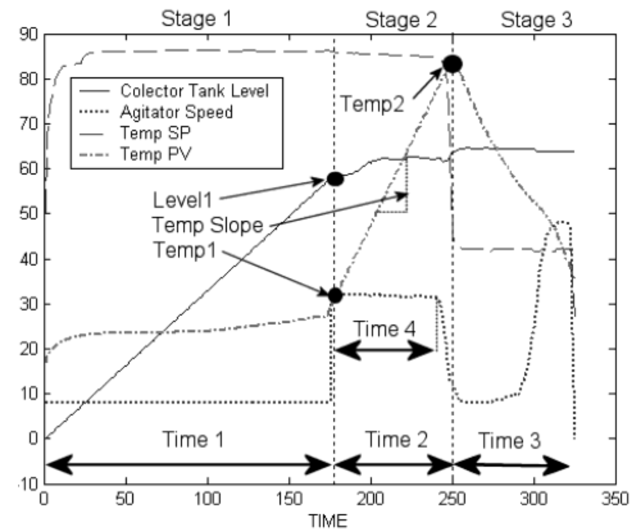

Figure 6. Stages and new variables defined during the batch (level in percent fill, temperature in degrees Celsius, speed in rpm).

increasing variable that always starts each batch at the same value and ends the batch at another given value, then alignment can often be achieved simply by resampling at constant intervals of the variable. Industrial examples of such indicator variables are conversion information[8] and the cumulative weight of a key monomer added during each batch.[16] If such an indicator variable does not exist over the whole duration of the batch, one may exist within each stage of the batch, allowing for a stage-by-stage alignment.

The discrete events that, most of the time, determine the transition between stages within a batch are triggered by operators or by the automation system in response to the achievement of certain conditions during the preceding stage. Kaistha and Moore[17] proposed a mathematical filter to extract the events from the batch trajectories. However, in practice we often already know the events, and there is no need to identify them from the trajectory.

In this industrial study, no one "key" monotonically increasing or decreasing indicator variable existed during the batch. Therefore, to align the batches, we used prior knowledge about how the process was operated to define three clear stages in the batch (Figure 6). Stage 1 runs from the beginning of the batch and up to the time where the agitator is turned to high speed; stage 2 runs from this point and up to the time where the temperature reaches its peak; and stage 3 is from this point until the end of the batch.

To align the batches, we use three different indicator variables (one per stage). For stage 1 the chosen indicator variable is the level in the collector tank. The value of this level at the end of the stage is very different for each batch, so it is not proper to simply resample using a fixed ΔLevel for all batches. Therefore, for any batch i, we assume that stage 1 is 0% complete when the tank is empty (time zero), and it is 100% complete when the level reaches the Level1_i value at the end of the stage; to achieve n samples between 0 and 100% of completion, each batch i is resampled at level increments given by $\Delta\text{Level}_i = \text{Level}_i/(n - 1)$, and linear interpolation is used when needed. Aligning this way ensures an equal number of samples, and a "line up" in the percentage of completion for all batches in stage 1.

For stage 2 the indicator variable is the dryer temperature. The alignment strategy is similar to that for stage 1 because the initial and final temperatures in this stage (Temp1 and Temp2) are very different for each batch. In this case stage 2 is completed 0% when the temperature is at its Temp1_i value and is completed 100% when the temperature reaches its peak value Temp2_i. For each batch we calculate a ΔTemp_i, and all variables (for stage 2) are resampled using this increment.

In stage 3 there is no variable that can be used as the indicator, so we resample linearly with time in order to obtain 75 samples total from the start of stage 3 to the end of the batch. However, the total time in stage 3 for each batch is recorded in the Z_w, and as discussed later in section 3.2, detailed warping information is also introduced as a new variable in the **X** matrix. In this way information on the duration of this stage and on the variation of the time usage throughout the batch is retained in the next PLS models developed in the next sections.

The alignment obtained can be appreciated by plotting some variables for all batches (Figure 7.) All batches now have 325 samples each, and all of the stages are synchronized, making them comparable. Figures 8 and 9 show the aligned trajectories of the two indicator variables used: the level in the collector tank and the dryer temperature.

With the definition of these stages, we can calculate eight new discrete variables per batch that defines the alignment information of the batches. These data are added to the **Z** matrix in Figure 2 and will be included in the analysis in the later sections:

1. Level1: the level in the collector tank at the end of stage 1.
2. Temp1: the temperature in the dryer at the end of stage 1.
3. Temp2: the temperature in the dryer at the end of stage 2 (peak).
4. Time1: the total length of stage 1 (time samples).
5. Time2: the total length of stage 2 (time samples).
6. Time3: the total length of stage 3 (time samples).
7. Time4: the total length of high-speed agitation (time samples).
8. TempSlope: the slope of the dryer temperature trajectory in stage 2.

3.2. Inclusion of "Time" as One Extra Trajectory per Batch. From the new **Z** variables obtained from the alignment, four of them have the "signature" on how the batch "used" time (Time1, Time2, Time3, and Time4). These are indeed four discrete values of time that summarize important information on timing differences among the batches. However, there is a more

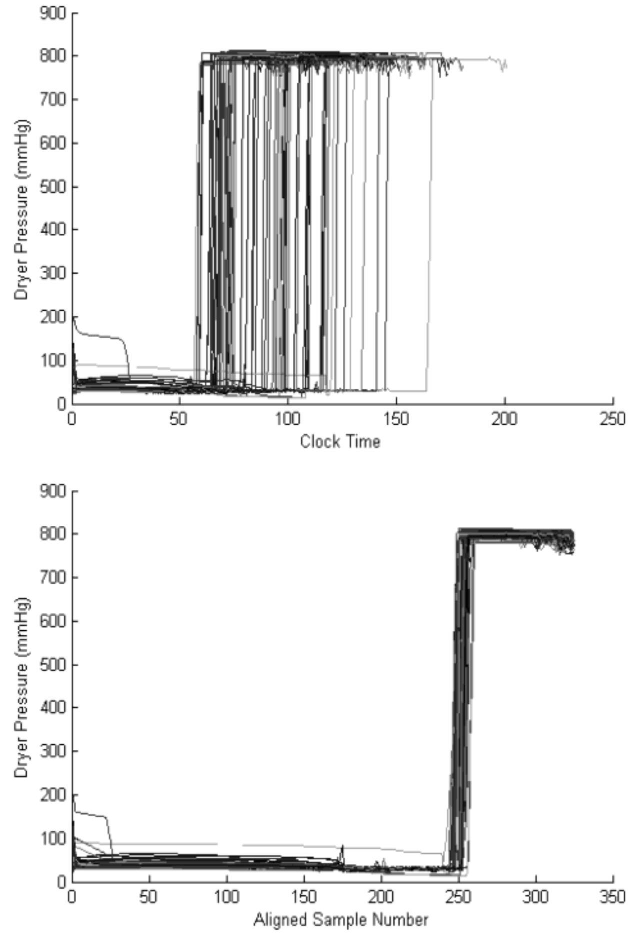

Figure 7. Variable 3 for all batches before (top) and after (bottom) alignment.

detailed way that we can extract timing information from the aligning exercise. In the original trajectories, time has a linear evolution for all batches. However, in the alignment process, this time is stretched and/or compressed as we resample. If we include "Time" as an extra variable for each batch and we resample this variable along with all of the others, at the end we will have a very useful trajectory that will tell us how each batch was "using" time as it evolved.[18] Therefore, batches which evolve faster or slower at different stages are readily apparent. Furthermore, if a fault is present in a particular batch, we will be able to say at what percentage of completion of the batch it happened.

Figure 10 shows how the "time" is distorted when we resampled the batch trajectories for this particular data set. These trajectories are added to the **X** matrix (as the 11th variable) and will be analyzed together with the other process variables in later sections.

4. Analysis of Historical Batch Data

The objective of this section is to uncover possible reasons for the occurrence of such a large number of batches that are producing off-spec products. PLS regression models are built to relate the final product quality (**Y**) to the aligned process trajectory data (**X**) and to the matrix (**Z**) containing the initial conditions for the batches (chemistry and charge) and the discrete timing variables. These PLS models project the information contained in the very large data matrices down into low-dimensional latent vector spaces, where it

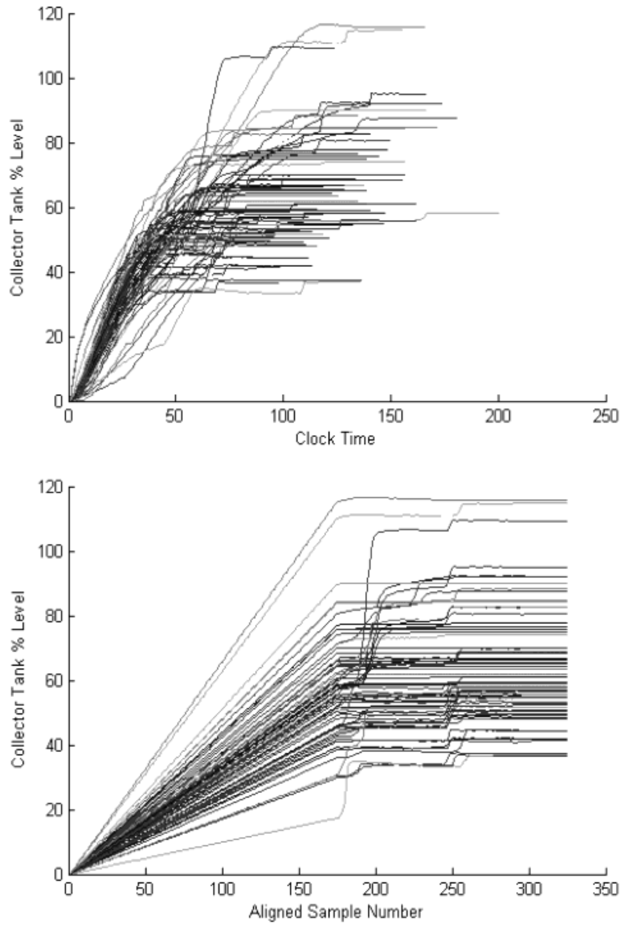

Figure 8. Collector tank level variable for all batches before (top) and after (bottom) alignment.

becomes much easier to visualize and compare the behavior of the batches and to diagnose the operational problems.

The block nature of the data (**Z**, **X**, **Y**) matrices available in this study naturally makes this problem suitable for analysis by multiblock PLS methods.[19,20] Multiblock PLS methods give the same prediction models for **Y** as a single-block PLS if the same weighting and scaling is applied to each variable[21] but additionally allow for estimation of the separate effect of each block. In this study, with only two regression blocks (**Z** and **X**), separate single-block PLS models relating **Z** to **Y** and **X** to **Y** were founded to provide essentially the same information as the multiblock approach. Therefore, for simplicity, only the former approach will be presented.

The analysis is presented in two sections: the first one is to understand the impact that the variables in the **Z** matrix have on quality and the second one to understand the impact that the trajectory variables in the **X** matrix have on quality.

4.1. Impact of the Initial Conditions and Warping Information (Z) on the Final Quality (Y). The **Z** matrix, at this point of the analysis, contains two different types of variables: (a) the information from the alignment and (b) the initial chemical and physical analysis. It was decided to partition the **Z** matrix into two blocks, Z_w and Z_{ch}.

Z_w is made with 9 columns of **Z** representing the variables arising from the alignment (or warping), plus the wet cake weight (Level1, Temp1, Temp2, Time1, Time2, Time3, Time4, TempSlope, and WgtCake), and Z_{ch} is made with the 11 columns of **Z** representing the chemical analyses (Z1, Z2, Z3, ..., Z11). It was decided to place the wet cake weight with the warping information because then all operating variables of the batch are together and separated from all of the chemical analyses.

The first step taken in this analysis is to remove from the data the 12 outliers found in section 2.2 and 15 more

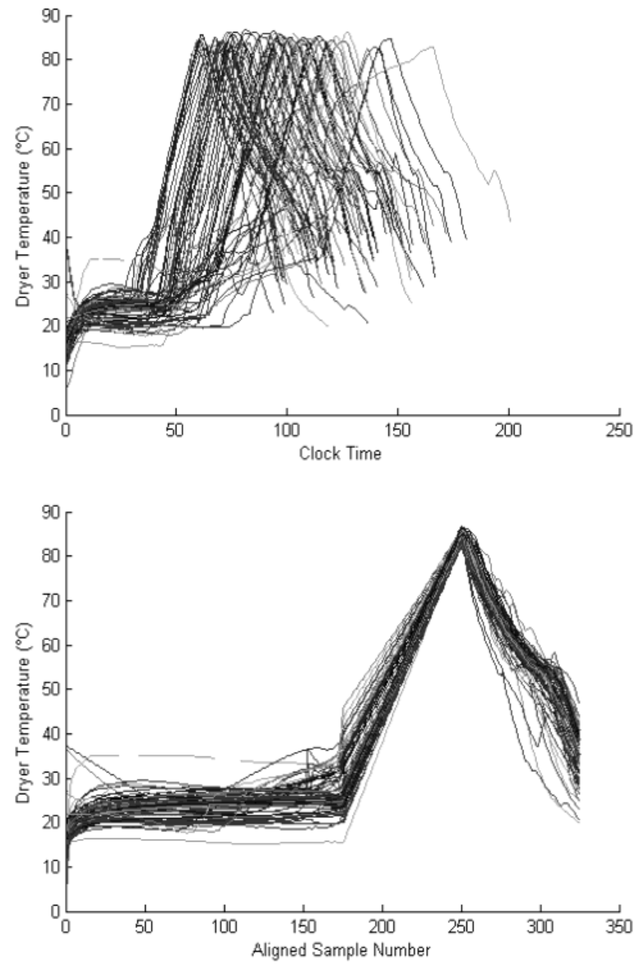

Figure 9. Dryer-temperature variable for all batches before (top) and after (bottom) alignment.

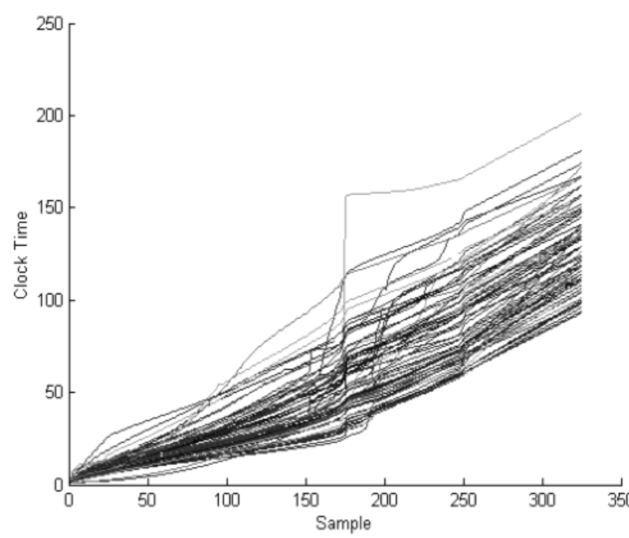

Figure 10. Distortion of clock time with the alignment of batch data.

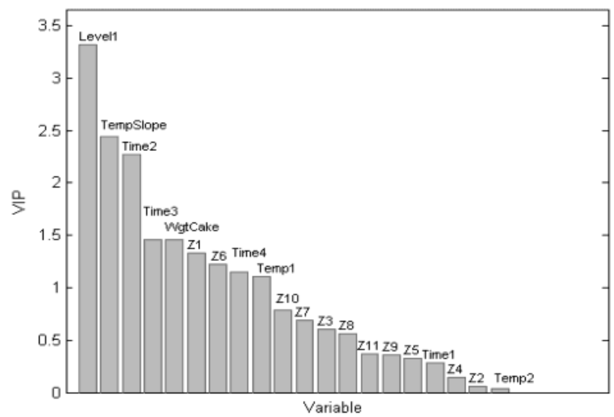

Figure 11. VIP plot for the **Z**−**Y** PLS model.

Table 2. Percentage Variance of Y Explained by Various PLS Models with Z

model	% variance of **Y** explained	model	% variance of **Y** explained
Z−**Y** PLS	50.49	Z_{ch}−**Y** PLS	21.52
Z_w−**Y** PLS	37.14		

observations for which the whole Z_{ch} information was missing. The final data set consists of 44 batches and 21 variables grouped in two blocks. These 44 batches are now used in the rest of the paper.

A PLS is fitted with the whole **Z** matrix, and for illustrative purposes, two more models are built individually for each block of **Z** (Z_w and Z_{ch}). The percentage of the variance that each model captures is listed on Table 2; all models are built using two latent variables, as determined by cross-validation. As expected, none of the individual models captures as much variance as they do together. The fact that the combined captured variance (50%) is less than the sum of both individual captured variances (37.14 + 21.52 = 58.66) is an indication that both blocks are not entirely orthogonal to each other. This is because some of the variables in the Z_w block are correlated with some of those in the Z_{ch} block. This analysis also shows that the process operating variables Z_w have a much greater impact on the product quality (**Y**) than the initial charge chemistry. This is an important result, as will be confirmed later.

The PLS of the whole **Z** matrix versus **Y** may be used to show the relative importance of each of the variables in the model. The VIP (variable importance to the projection)[22] plot in Figure 11 lists the variables in **Z** in their order of importance. Several important conclusions arise from this VIP plot: (i) it confirms the previous conclusion that the most important variables are those variables related with the operation itself and not those related to the initial chemistry; (ii) the most important variable is the initial amount of solvent in the wet cake; (iii) the variable Temp2 appears to be the least important of all of the variables, although this variable was the one "adjusted" in the plant in order to correct the quality. Because the manipulated variable is involved in some sort of feedback loop to control some aspect of the final quality, its relationship with **Y** will be some combination of the open-loop effect and the closed-loop effect on **Y**, and such a result is not surprising.

Additional important results can be obtained by looking at the score space of this model (Figure 12). It

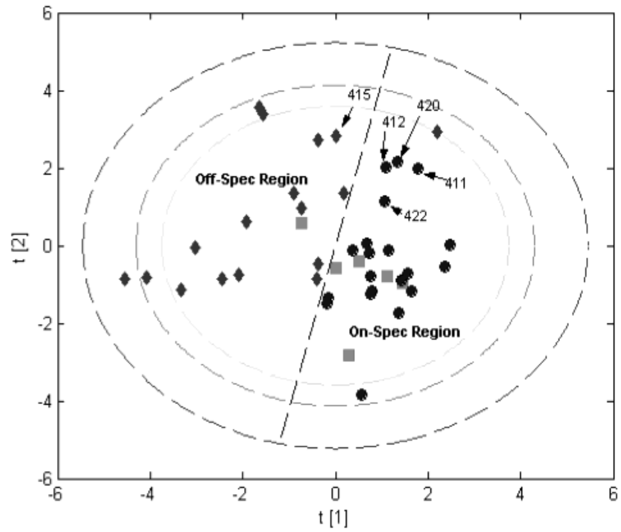

Figure 12. Score plot for the **Z**−**Y** PLS model: off-spec (◆), on-spec (●), on-spec but high residual solvent (■) (numbered batches are references for analysis in section 4.3).

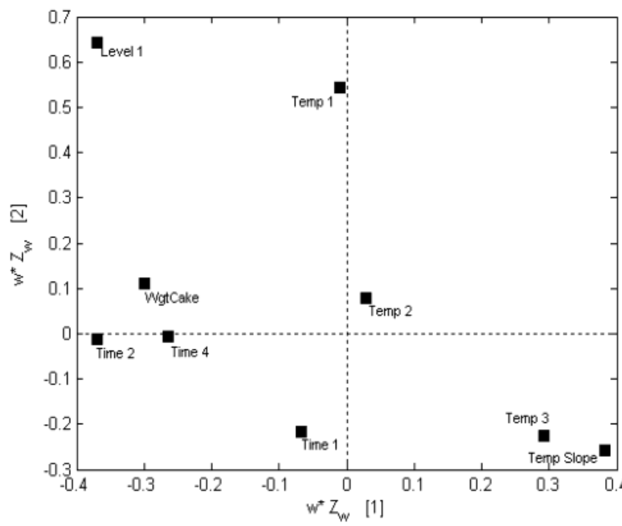

Figure 13. Loadings plot for the Z_w variables in the **Z**−**Y** PLS model.

is clear from this plot that most of the good-quality batches have a positive value of t_1, while the poor-quality ones have a negative value of t_1.

If the loadings for Z_w (Figure 13) and Z_{ch} (Figure 14) are now analyzed in light of this new piece of information (good-quality batches have high positive values of t_1), we can tell much more about the process.

From the loadings map for the Z_w variables (Figure 13), it is clear that a high-quality product is associated with a high temperature slope in stage 2 and low values of the length of stage 2 (Time2) and low high-speed agitation time (Time4). All of these imply that fast evaporation at stage 2 is desirable. A high-quality product is also associated with low solvent level in the collector tank at the end of stage 1 (Level1) and a low initial charge of wet cake (WgtCake), both implying that a smaller charge of wet cake (and hence low solvent) is desirable. A slow cooldown in stage 3 (Time3) and possibly a low Temp1 might also seem desirable.

From the loadings for the Z_{ch} variables (Figure 14), it also appears that high quality might also be related to higher levels of chemicals Z6 and Z7, and lower levels of Z1, in the wet cake.

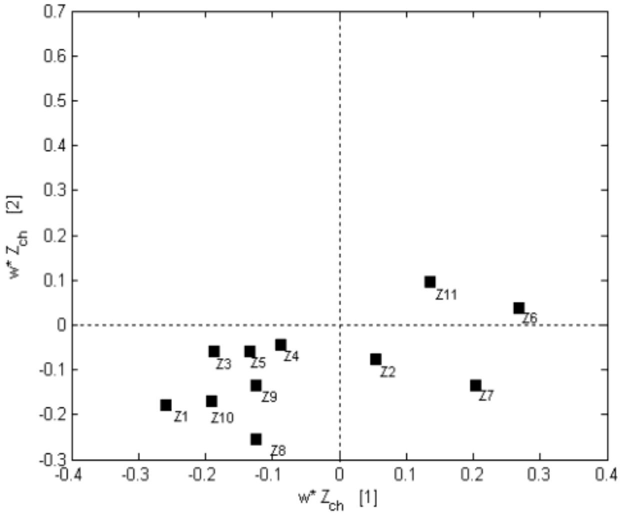

Figure 14. Loadings plot for the Z_{ch} variables in the **Z**−**Y** PLS model.

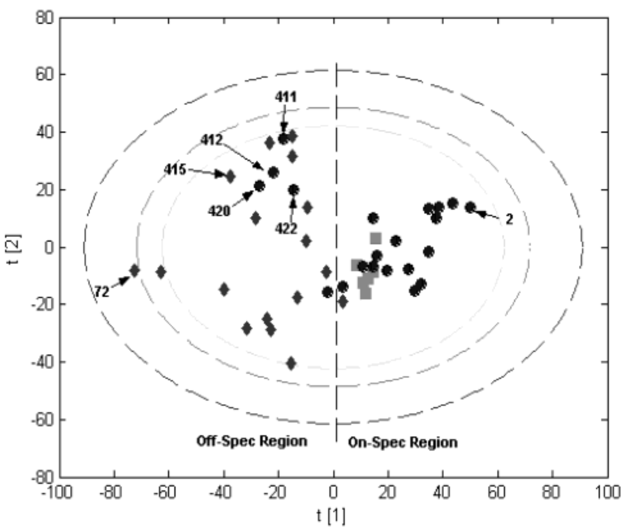

Figure 15. Score plot for the **X** space in the **X**−**Y** MPLS: off-spec (◆), on-spec (● and ■) (numbered batches are references for analysis in section 4.3).

4.2. Impact of the Process Variable Trajectories on the Final Quality. This section covers the analysis on how the shape and timing of the process variable trajectories (**X**) affect the final quality. It is important to keep in mind that the variables are evolving with time and the effect of a variable at different points of times is not the same. The analysis will not be done on single variables but on their trajectories summarized in the **X** matrix.

The **X** matrix has 11 variables—10 process variables and the clock time from the alignment—each with 325 samples per batch, giving a final unfolded **X** matrix having 3575 columns (variables at all time intervals) and 44 rows (batches). The multiway PLS model captures 36% of the variance of the **X** matrix and explains 48% of the **Y** matrix with 2 latent variables.

The score plot for the **X** space is shown in Figure 15. Most of the batches with similar quality do cluster together except for batches 411, 412, 420, and 422; these batches will be discussed later in section 4.3. Before doing so, we need to interpret the model in order to better understand the position of the batches in the score plot.

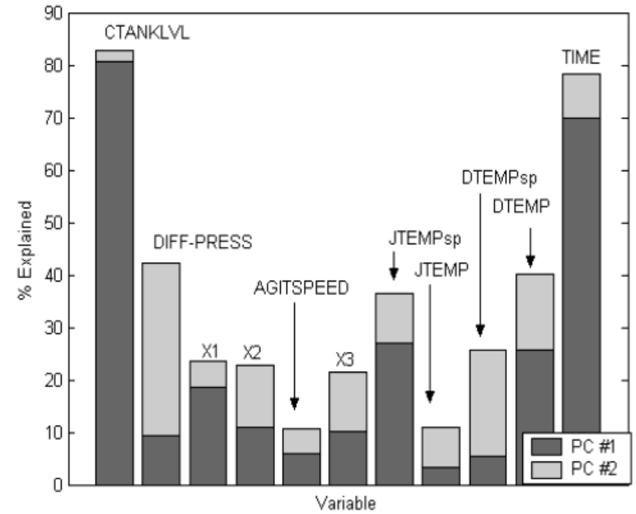

Figure 16. Percentage captured per variable in the **X** matrix.

Interpretation of the Multiway PLS Model. (a) Captured Variance. The way the model is capturing the variable trajectory information can be analyzed in different ways because each variable in the **X** matrix varies with time and has a different importance in each time interval. Although the model uses only 36% of the information in the **X**, this is the average for all variables at all time samples: Figure 16 shows the total percentage explained per variable per component; the variables CTANKLVL, DIFF-PRESS, JTEMPsp, DTEMP, and TIME are variables that are used much more fully by the model and hence are the most important ones in describing quality. Notice the high importance of the time usage; this re-enforces our conclusion of the past section about the importance of timing. To better understand the importance of each of the process variables over their time history, we will analyze the weight of each variable trajectory as it changes with time.

(b) Analysis of the PLS Model Weights. Because each variable in the **X** matrix is a trajectory, their loadings (or weights in the case of PLS) are not single values but are trajectories indicating the importance of each variable at each point of time to the latent variables. If, for the first component, we plot the weight trajectory for variable 1 from sample 1 to sample 325 and beside it we plot the weight trajectory for variable 2 from sample 1 to sample 325 and we continue in a similar manner for all variables, we end up with Figure 17. The magnitude of the weight, for a certain variable, at a certain point in time is an indication of how important that variable is, at that point in time, for the first component (t_1).

It is clear from the score plot in Figure 15 that a high positive value of t_1 leads to high-quality products. With this in mind, we can therefore interpret the trajectory loadings in Figure 17.

A high-quality batch will have the following characteristics:

1. It will have low levels of the solvent collector tank level (CTANKLVL) with respect to its mean trajectory from the model, throughout the entire batch. This result clearly is consistent with the importance found previously in the PLS analysis of the **Z** variables that a low wet cake fed to the dryer (WgtCake) and a low level in the collector tank at the end of stage 1 (Level1) are key factors for on-spec product quality.

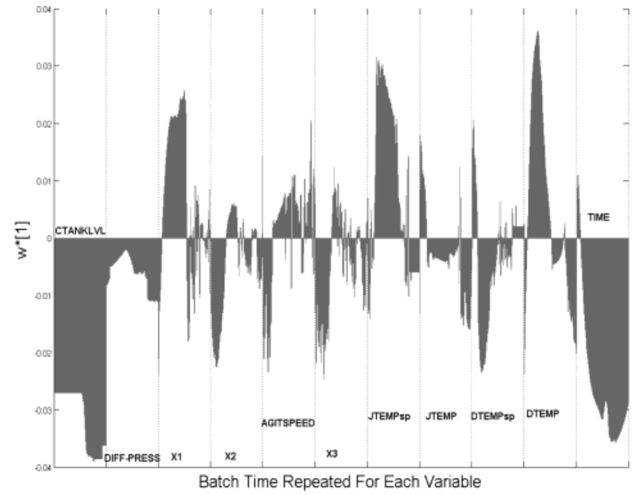

Figure 17. Weights for the first component in the **X** space for the **X**−**Y** MPLS.

2. It will have high levels of pressure in the dryer (X1), the set point for the jacket heating medium (JTEMPsp), and the temperature inside the dryer (DTEMP) with respect to each of their mean trajectories from the model for the first stage of the batch. These all appear to stress the importance of applying a high level of heating during stage 1.

3. It will have low values of the time variable (TIME) relative to its mean trajectory from the model throughout most of the batch. This implies that batches that progress faster tend to be those with high product quality.

To illustrate this, consider two batches on opposite sides of the score plot in the t_1 dimension (Figure 15): batch 72 (which has a high negative value of t_1) and batch 2 (which has a high positive value of t_1). For these batches we plot the collector tank level (CTANKLVL) for each of them along with the mean trajectory (Figure 18, top), and we see that the trajectory for this variable in batch 2 is always below the mean as opposed to batch 70. If we plot the temperature DTEMP for these batches (Figure 18, bottom), we see that batch 2 has a value of DTEMP above the mean for the first stage and below the mean for the second and third stages of the batch, as opposed to batch 70 which has a value below the mean for the first stage and above the mean for the second and third stages of the batch.

The above analysis performed on the variable trajectory weighting for the t_1 variable clearly provided insights into the operating policies that appear to lead to a good- or poor-quality product.

4.3. Anomalous Batches 411, 412, 420, and 422. In Figure 15 four batches were observed to lie in the off-spec region of the t_1-t_2 score plot for the **X**−**Y** MPLS model. These same four batches clustered in the on-spec region in the **Z**−**Y** PLS score plot (Figure 12) and eventually yielded good product at the end. The fact that these four batches cluster in the off-spec region in Figure 12 means that the process variable trajectories of these batches have the characteristics of other "bad" batches. However, it appears that some conditions in the **Z** matrix were sufficient to counterbalance these bad trajectories.

To reassure the fact that the trajectories of these four batches have the same characteristics of any other "bad" batch, we compared (looking for similarities) the con-

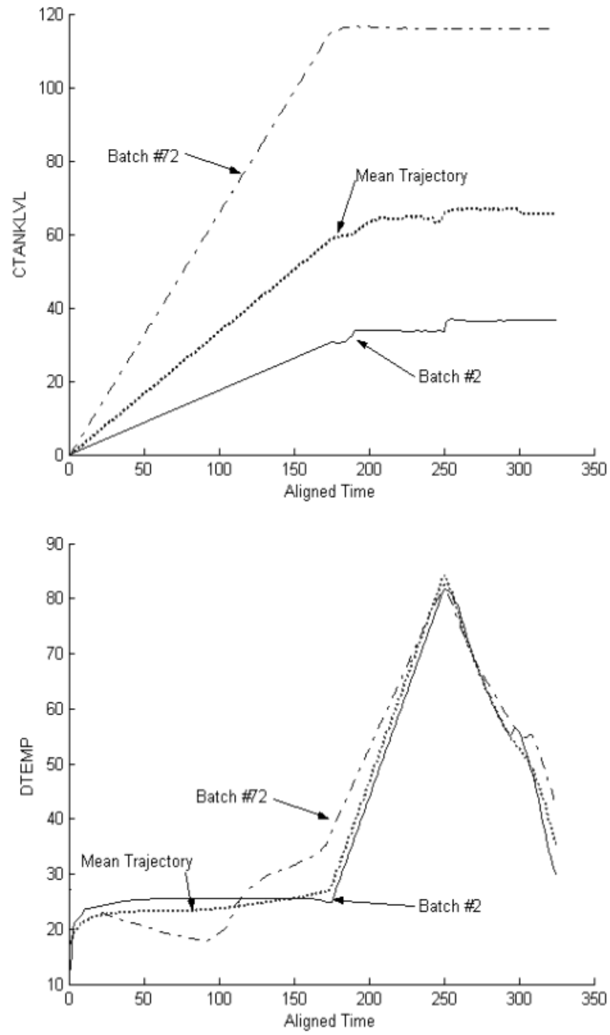

Figure 18. Differences in the trajectory of the collector tank level between batches 72 and 2.

tributions plots[23−25] in t_1 (using the **X**−**Y** PLS model) for these four batches with the contribution plots in t_1 of other "bad" batches in the vicinity of these batches. $-t_1$ is used because this component defines the on-spec/off-spec division, as illustrated in Figure 15.

Figure 19 shows the contribution plots for t_1 for batches 420 (anomalous batch) and 415 (bad batch) in the **X**−**Y** MPLS model. Both plots show the same characteristics: the trajectory for the collector tank level (CTANKLVL) is above the mean during the entire trajectory; the variables X1, JTEMPsp, and DTEMP are all below their mean during the first stage, implying too slow heating and this contributes to a long CLOCK-TIME. All of these conditions are totally contrary to the ones related with good on-spec quality product, according to our earlier conclusions. These contributions are very similar to other nearby batches in the score plot (like batches 411, 412, and 422).

Now consider the **Z**−**Y** PLS model score plot (Figure 12). Most of the "bad" batches that are near the anomalous ones in the **X**−**Y** score space (Figure 15) appear apart in the **Z**−**Y** score space (Figure 12), e.g., batches 415 and 420.

Figure 20 is a contribution plot for t_1 in the score space for the **Z**−**Y** PLS model, *from* batch 415 *to* batch 420. This contribution plot shows that batch 420 had a steeper slope for the temperature in stage 2 (high value of TempSlope and low value of Time2) and a much

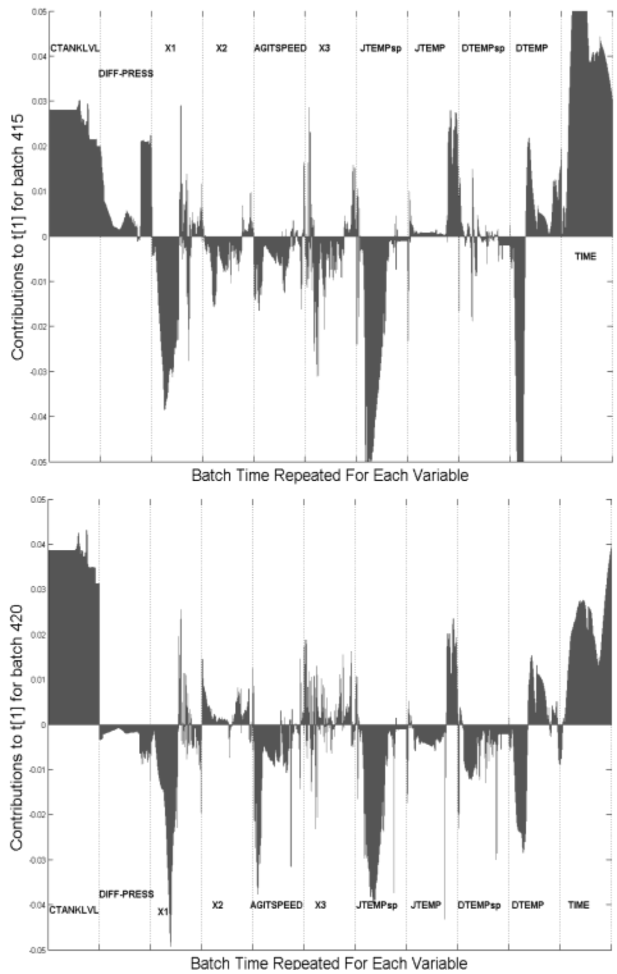

Figure 19. Contribution plots to t_1 in the **X**−**Y** MPLS model for batch 415 (top) and 420 (bottom).

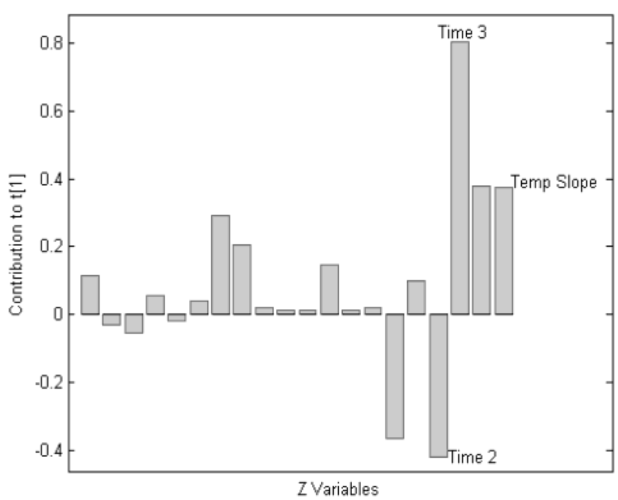

Figure 20. Contribution plots to the scores t_1 in the **Z**−**Y** PLS model from batch 415 to batch 420.

larger cooldown time (Time3). These two characteristics (high evaporation speed in stage 2 and large cooldown time) also appear in all of the contribution plots between the other anomalous batches and the bad batches, indicating that a fast evaporation in stage 2 and a large cooldown time are “strong” enough characteristics to compensate for all of the other adverse conditions in the trajectories and made these batches yield a good product at the end.

5. Conclusions

Multivariate statistical methods were used to analyze historical data from an industrial batch drying process to uncover possible causes for a high level of poor-quality product. The methods were able to show that the initial chemistry of the charge had little impact on the quality and that it was variations in the operating policies of the batches that were the dominant contributions to the poor-quality product. These methods were further able to isolate individual variables such as the initial wet cake charge, timing variables such as the duration of the second and third stages, and process variable trajectories such as the temperature and pressure profiles during the first stage of the batch that were highly related to poor final product quality. This study illustrates the potential of these multivariate methods for analyzing historical batch data and for suggesting process improvements. Some new approaches to the alignment of the batch trajectories and the subsequent use of the alignment information in the analysis were introduced.

6. Software Used

All mathematical calculations presented in this work were done using MACSTATv4 and BatchSPCv2; both are MATLAB toolboxes developed at the McMaster Advanced Control Consortium.

Acknowledgment

S.G.M. thanks the McMaster Advance Control Consortium for its funding. The authors thank FMC Corp. for allowing them to use the data from the process.

Literature Cited

(1) MacGregor, J. F.; Nomikos, P. Monitoring Batch Processes. *NATO Advance Study Institute for Batch Processing Systems Engineering*, Antalya, Turkey, 1992; Springer-Verlag: Heidelberg, Germany, 1992.

(2) Nomikos, P.; MacGregor, J. F. Monitoring Batch Processes using Multiway Principal Components. *AIChE J.* **1994**, *40* (8), 1361−1375.

(3) Nomikos, P.; MacGregor, J. F. Multivariate SPC Charts for Monitoring Batch Processes. *Technometrics* **1995**, *37* (1), 41−58.

(4) Nomikos, P.; MacGregor, J. F. Multi-way partial least squares in monitoring batch processes. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 97−108.

(5) Wold, S.; Geladi, P.; Esbensen, K.; Ohman, J. Multi-way Principal Components and PLS analysis. *J. Chemom.* **1987**, *1*, 41−56.

(6) Kourti, T.; Nomikos, P.; MacGregor, J. F. Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS. *J. Process Control* **1995**, *5* (4), 277−284.

(7) Kosanovich, K.; Dahl, K. S.; Piovoso, M. J. Improved Process Understanding Using Multiway Principal Component Analysis. *Ind. Eng. Chem. Res.* **1996**, *35*, 138−146.

(8) Neogi, D.; Schlags, C. E. Mutivariate Statistical Analysis of an Emulsion Batch Process. *Ind. Eng. Chem. Res.* **1998**, *37*, 3971−3979.

(9) Schlags, C. E.; Popule, M. P. Industrial Application of SPC to Batch Polymerization Process. *Proc. ACC (Arlington, VA)* **2001**, 25−27.

(10) Nomikos, P. Detection and diagnosis of abnormal batch operations based on multi-way principal component analysis. *ISA Trans.* **1996**, *35* (3), 259−266.

(11) Tates, A. A.; Louwerse, D. J.; Smilde, A. K.; Koot, G. L. M.; Berndt, H. Monitoring a PVC batch process with multivariate statistical control charts. *Ind. Eng. Chem. Res.* **1999**, *38*, 4769−4776.

(12) Wold, S.; Kettaneh, N.; Friden, H.; Holmberg, A. Modelling and diagnostics of batch process and analogous kinetic experiments. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 331−340.

(13) Boque, R.; Smilde, A. K. Monitoring and Diagnosing Batch Processes with Multivariate Covariates Regression Models. *AIChE J.* **1999**, *45* (7), 1504−1520.

(14) Duchesne, C.; MacGregor, J. F. Establishing Multivariate Specification Regions for Incoming Materials. *J. Qual. Technol.*, submitted for publication.

(15) Kassidas, A.; MacGregor, J. F.; Taylor, P. Synchronization of batch trajectories using dynamic time warping. *AIChE J.* **1998**, *44* (4), 864−875.

(16) Kourti, T.; Lee, J.; MacGregor, J. F. Experiences with Industrial Applications of Projection Methods for Multivariate Statistical Process Control. *Comput. Chem. Eng.* **1996**, *20*, S745−S750.

(17) Kaistha, N.; Moore, Ch. Extraction of Event Times in Batch Profiles for Time Synchronization and Quality Predictions. *Ind. Eng. Chem. Res.* **2001**, *40*, 252−260.

(18) Westerhuis, J. A.; Kassidas, A.; Kourti, T.; Taylor, P. A.; MacGregor, J. F. On-line synchronization of the trajectories of process variables for monitoring batch processes with varying duration. Presented at the 6th Scandinavian Symposium on Chemometrics, Porsgrunn, Norway, Aug 15−19, 1999 (submitted to *AIChE J.*).

(19) Wangen, L. E.; Kowalski, B. A Multiblock Partial Least Squares Algorithm for Investigating Complex Chemical Systems. *J. Chemom.* **1988**, *3*, 3−20.

(20) Westerhuis, J. A.; Smilde, A. K. Short Communication: Deflation in multiblock PLS. *J. Chemom.* **2001**, *15*, 485−493.

(21) Westerhuis, J.; Kourti, T.; MacGregor, J. F. Analysis of Multiblock and Hierarchical PCA and PLS Models. *J. Chemom.* **1998**, *12*, 301−321.

(22) Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wold, S. *Multi- and Megavariate Data Analysis: Principles and Applications*; Umetrics: Umea, 1999.

(23) Miller, P.; Swanson, R. E.; Heckler, C. F. Contribution Plots: The Missing Link in Multivariate Quality Control. *Appl. Math. Comput. Sci.* **1998**, *8*, 775−782.

(24) MacGregor, J. F.; Jaeckle, C.; Kiparissides, C.; Koutoudi, M. Process Monitoring and Diagnosis by Multiblock PLS Methods. *AIChE J.* **1994**, *40* (5), 826−838.

(25) Kourti, T.; MacGregor, J. F. Multivariate SPC Methods for Process and Product Monitoring. *J. Qual. Technol. 28*, 9−428.

Received for review January 7, 2003
Revised manuscript received May 2, 2003
Accepted May 14, 2003

IE0300023