

Data Analytics Can Help Reduce Energy Consumption in the Industrial Manufacturing of Specialty Chemicals

Francesco Sartori^a, Federico Zuecco^b, Pierantonio Facco^a, Fabrizio Bezzo^a,
Massimiliano Barolo^{a,*}

^aCAPE-Lab – Computer-Aided Process Engineering Laboratory, Department of Industrial Engineering, University of Padova, 35131 Padova PD (Italy)

^bBASF Italia SpA, Via Pila 6/3, 40037 Pontecchio Marconi BO (Italy)
max.barolo@unipd.it

Batch processes are operated following recipes that consist of a sequence of steps of given time lengths carried out in different pieces of equipment. Large variability in the length of a processing step can cause that step to become a bottleneck for the entire process, thus leading to an increase of the energy consumption per unit of product manufactured. Debottlenecking the process can therefore lead to reduction of the energy requirements. We consider the case of a batch reaction that is a key step in the industrial manufacturing of a polymer additive. The available data historians revealed that, over a period of 12 months of operation, the length of the reaction step ranged between 0.9 and 2.8 h, with an average value of 1.3 h. This acted as a limit to the performance of the overall manufacturing system, but no cause was initially identified to explain this behavior. Advanced analytics on the process data historians by means of multivariate statistical techniques revealed that over 40% of the batches had been affected by intervention of a safety interlock in the reactor, whose occurrence strongly correlated to an increase of the batch length. Reconfiguration of the interlock system resulted in a reduction of both average batch length and batch length variability. Namely, over the 6-month assessment that followed this study, a 29% reduction in the average batch length for the reactor under investigation was observed, which resulted in an 8% reduction of the overall process cycle duration, thus entailing significant energy savings. Furthermore, an 11% reduction on nitrogen consumption was achieved.

1. Introduction

Batch processes are widespread in many industries producing low volumes of high added-value goods such as pharmaceuticals, biotechnological products and specialty chemicals. The energy costs for common chemicals produced in batch operations can arrive at as high as 10% of the total production costs (Bieler et al., 2004). Therefore, reducing energy consumption not only reduces the environmental impact of the process, but can also significantly reduce the process operating expenses.

One approach for modeling batch processes is based on first-principles models, requiring detailed knowledge about the phenomena occurring in the process. The development of these models is usually expensive, time consuming, hence often prohibitive in an industrial setting. On the other hand, the increased availability of data in the process industry, propelled by the development of sensors and networking technology together with the reduction of the costs of computing equipment, allowed the development of data-driven models for tasks traditionally carried out through first-principles models, thus sensibly reducing time and costs for model development. Batch processing is highly impacted by this approach, especially when the process chemistry is not completely understood, which renders the development of a first-principles model a hard (or even impossible) task.

In order to extract process-relevant information from the massive amount of data generated by a modern chemical process, effective data analytics techniques can be used. Multivariate statistical methods, such as principal component analysis (PCA; Jolliffe and Cadima, 2016) and its multiway extension (Nomikos and MacGregor, 1995) are extensively used to this purpose. These techniques can reduce the dimensionality of

large sets of data, thus increasing their interpretability while minimizing information loss, and revealing the underlying correlation structure between the process variables over their time evolution. They do this by projecting the data onto the space of a reduced set of new, uncorrelated variables (called principal components) that summarize the original data, in such a way that an intuitive visual comparison of the data evolution patterns across different batches can be obtained.

In this study, we exploit PCA to find the root-cause determining a large variability in the time duration of a key reaction step for an industrial batch process manufacturing a specialty chemical. Large (and unexplained) average batch length and length variability in this reaction step caused significant energy and raw materials consumption per unit of product manufactured, yet with no apparent impact on the quality of the final product.

2. Process description

The process under investigation consists in the synthesis of an intermediate chemical for the manufacturing of a hindered amine light stabilizer (HALS), to be used as a polymer additive. A simplified process flow diagram is shown in Figure 1.

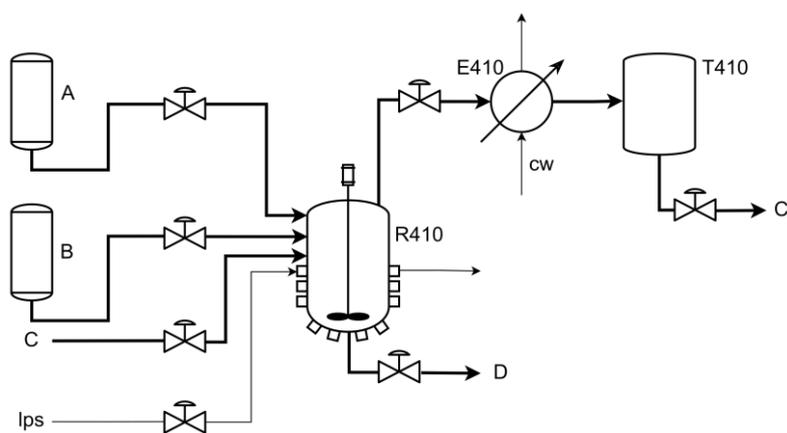


Figure 1: Simplified process flow diagram of the process under investigation

In this semi-batch process, reactants A and B are fed to the jacketed stirred tank reactor R410. Vacuum is then applied to the system, and the main reaction is a liquid-phase, thermally activated, exothermic one, according to:



where D is the desired product, and C is a byproduct.

The process is highly automated and is operated through a recipe that consists of the following steps:

1. setup: the system is set up for a new run;
2. reactant loading: liquid reactants A and B are loaded into R410 from their respective storage tanks in stoichiometric amounts. A small amount of C is also loaded into the reactor; the reason for this is to speed up the initial reaction phase. The reactor is heated up with low-pressure steam through a jacket, to reach the temperature required to carry out the reaction;
3. reaction: vacuum is applied to the system in two steps: a faster pressure decrease is applied first, down to an assigned pressure; then, pressure is further slowly decreased to the pressure value required by the reaction. Once the reaction conditions are met, byproduct C is released as a vapor, and it is removed from the reactor, condensed in E410 condenser, and then stored in T410 buffer tank;
4. nitrogen blanketing: when the required volume in T410 is obtained, vacuum is broken, and the reactor is blanketed with nitrogen;
5. product discharge: liquid byproduct C is discharged from T410 to the waste unit, reaction product D (liquid) is discharged from R410 and fed to the subsequent processing step.

Step 3 corresponds to the reaction phase (the key one for this process), and we call "batch length" its duration. Figure 2 shows the distribution of batch lengths recorded over a period of 12 consecutive months before this study was started. It can be seen that the distribution is bimodal, with one mode with a peak at 56 min and one mode with a peak at 74 min; furthermore, the batch lengths range between 55 and 145 min, and the overall average batch length is 77 min.

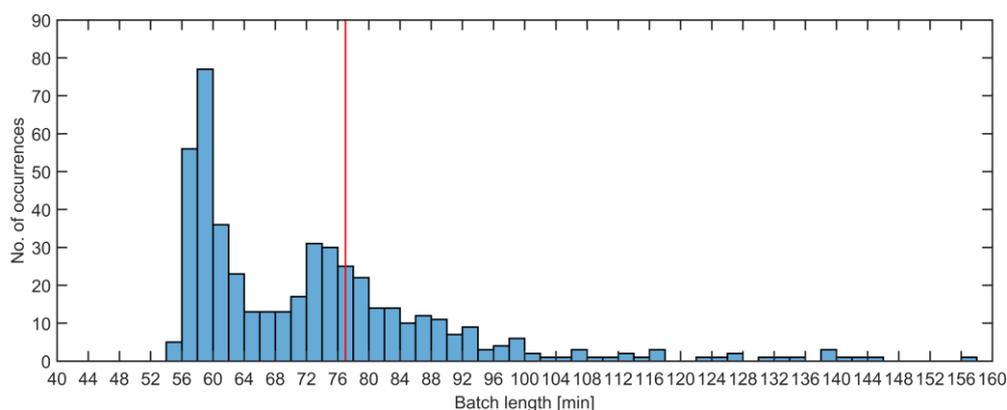


Figure 2: Distribution of the time duration of step 3 in reactor R410 across the historical dataset. The vertical red line is the mean of the distribution.

The result of this variability in batch length is a decrease in productivity, as well as an increase in energy consumption per unit of product manufactured. Since engineering understanding was not enough to find the root cause of this variability, analytics on the historical manufacturing data was done to mine process-relevant information that could help in the task of troubleshooting the reaction step.

3. Mathematical background

3.1 Principal component analysis

A short introduction on principal component analysis (PCA) is provided in this section. Detailed information can be found elsewhere (Jolliffe and Cadima, 2016). Let $\mathbf{X}[I \times J]$ be a historical dataset composed of I observations and J variables. PCA is a multivariate statistical technique that allows describing \mathbf{X} through $A \leq \min(I, J)$ principal components. In order to determine the most appropriate value of A for a given dataset, the scree test can be used (Brown, 2009).

PCA models the historical dataset as:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (2)$$

where $\mathbf{T}[I \times A]$ is the scores matrix, $\mathbf{P}[A \times J]$ is the loadings matrix, and $\mathbf{E}[I \times J]$ is the residuals matrix.

In batch processes, measurements of the time trajectories for several process variable are collected continuously in time during each run. A convenient arrangement for batch process data is therefore a tensor-like one, namely $\underline{\mathbf{X}}[I \times J \times K]$, where $\underline{\mathbf{X}}$ is made by I batch runs, J measured variables, and K time samples for them. In order to analyze a three-way array like $\underline{\mathbf{X}}$, multiway PCA (MPCA) can be used. This technique is applied by unfolding $\underline{\mathbf{X}}$ along its second dimension (i.e., batchwise unfolding; Camacho et al., 2008), thus obtaining $\mathbf{X}_U [I \times JK]$, and then applying standard PCA to this two-way matrix (Nomikos and MacGregor, 1995).

3.2 Batch alignment

In recipe-driven, multiphase batch processes, such as the one considered in this study, the termination of each phase is usually triggered by an event that indicates phase completion (e.g., reaching a desired volume of separated byproduct). Often, these events occur at a different time in each batch, and each batch may therefore come to completion at a different time (Undey and Cinar, 2002). In order for multivariate statistical techniques to be applied to data coming from processes with these uneven-length characteristics, batch alignment is required. In this study, the indicator variable batch alignment technique was applied (Nomikos and MacGregor 1994). This technique assumes that a variable exists in the dataset that is a reliable indicator of the process evolution. A variable is a good candidate as an indicator variable if (García-Muñoz et al., 2003):

- it is monotonic in time;
- it has a favorable signal-to-noise ratio;
- it has the same initial and final values across all batches.

For a multiphase process, each phase can be aligned using a different indicator variable (García-Muñoz et al., 2003). When such a variable is not present in the dataset, time can be used as the indicator variable, either for the whole process, or for each single phase.

4. Available data

A total of $I = 468$ historical batches were extracted from the plant historian (Figure 2). All batches ended up in a product meeting the target quality profile. The available data consists of the time trajectories of 7 operating variables, as listed in Table 1. The measurements were collected every 30 s. The available data were organized in a tensor $\underline{\mathbf{X}}[I \times J \times \bar{K}]$, where $J = 7$ is the number of measurement sensors available for the unit, and \bar{K} is the total number of observations per batch, ranging between 183 and 1225.

Table 1: Available variables in the $\underline{\mathbf{X}}$ dataset

Variable no.	Variable description
1	R410 absolute internal pressure
2	R410 internal pressure controller output
3	R410 internal temperature
4	E410 condensed liquid temperature
5	T410 internal level
6	R410 internal level
7	Time

5. Analysis of historical batch data

The dataset was aligned with the indicator variable technique applied to each operating phase, using the internal level as the indicator variable for phases 2 and 5, and time as the indicator variable for phase 3, thus obtaining $\underline{\mathbf{X}}_a[468 \times 7 \times 296]$. Due to their very short time duration, operating stage 1 and 4 were neglected. An MPCA model with 3 PCs was calibrated utilizing the preprocessed dataset $\underline{\mathbf{X}}_a$. The model explains 51% of the dataset variability.

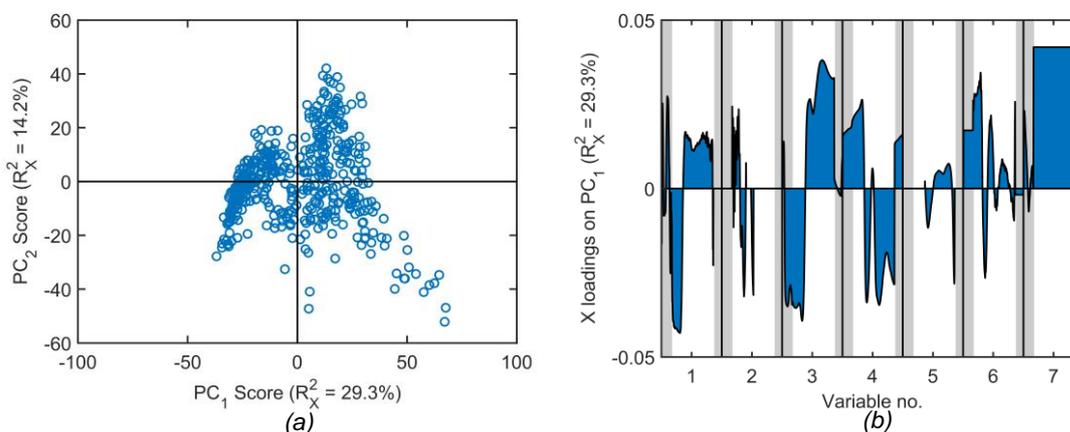


Figure 3. Results of the multiway principal component analysis (MPCA) model built on the historical process data: (a) scores plot, (b) loadings plot on PC1 (time evolution for each process variable). The grey areas shown for each variable in (b) refer to step 2 (left) and step 5 (right) of the manufacturing recipe; the white area refers to step 3.

The model scores for the first two PCs are plotted in Figure 3a: no particular pattern across the historical batches is apparent. On the other hand, analysis of the loadings (Figure 3b) reveals interesting information. The loadings plot shows how the loadings along the first PC evolve with the aligned time for each of the variables listed in Table 1. For any given variable, the time evolution is marked using three background colors: *i*) a grey left area, corresponding to the reactant loading phase (step 2); *ii*) a white central area, corresponding to the reaction phase (step 3); and *iii*) a grey right area, corresponding to the product discharge phase (step 5). Paired analysis of the loadings and scores plot (Bro and Smilde 2014) helps understanding how each variable concurs to separating the scores along a particular direction in the scores plot. For example, considering the direction along the first PC in the scores plane (left to right in Figure 3a), we can conclude from Figure 3b that the batches located in the right-half plane of Figure 3a are characterized by:

- lower pressure in the first part of the reaction phase, and higher pressure in the second part of the reaction stage (variable no. 1);

- lower temperature in the first part of the reaction stage, and higher temperature in the second part of the reaction stage (variable no. 3);
- lower temperature of the condensed liquid in the condenser in the reaction stage (variable no. 4);
- consistently longer duration of the reaction stage (variable no. 7).

Figure 4a reports the time profiles of the reactor internal pressure for two historical batches, respectively projecting onto the left-half plane (batch no. 18) and the right-half plane (batch no. 337) of the scores plot.

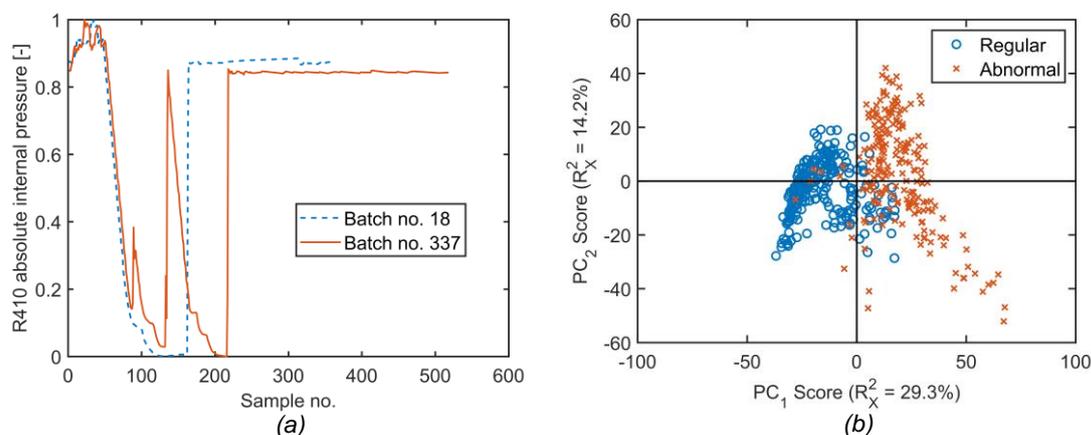


Figure 4: (a) Internal pressure profiles from a batch projecting onto the left-half plane of Fig. 4a (batch no.18) and a batch projecting onto the right-half plane of Fig. 4a (batch no.337). (b) MPCA model scores of Figure 3a colored according to whether they are designated as regular or abnormal.

It is observed that the internal pressure for batch no.18 follows the recipe described in Section 2 correctly, whereas batch no.337 shows an abnormal pressure profile, with two spikes in the first quarter of the batch duration. This indicates that vacuum was broken (and then reinstated) in that batch before reaching the end of the reaction stage. Confirming the conclusions drawn from the loadings analysis, batch no.337 have a longer duration than batch no.18 (Figure 4a), and higher internal pressure during the reaction phase (namely, in-between the initial decreasing ramp and the final step increase). Since loss of vacuum is an abnormal event (yet not leading to abnormal product quality), an algorithm to automatically identify all historical batches with a reactor pressure profile qualitatively similar to the one of batch no.337 was developed. The relevant batches were denoted as “abnormal”, to distinguish them from the “regular” ones, where the vacuum breakage event did not occur. It was found that as many as 40% of the historical batches can be designated as abnormal. The regular and abnormal batches were then identified in the scores plot, obtaining the results of Figure 4b. We notice that most of the abnormal batches are projected onto the right-half plane, whereas most of the regular ones are lying in the left-half plane, i.e., the separation between regular and abnormal batches occurs along the first PC. Therefore, we conclude that the difference between regular and abnormal batches acts as the strongest source of variability within the historical dataset.

Discussion with the process experts revealed that the abnormal pressure profiles observed in the historical dataset are related to the intervention of the reactor safety interlock system. In fact, the occurrence of particular combinations of operating conditions in the reactor can trigger the intervention of specific interlocks, each of which acting by breaking the vacuum and blanketing the reactor with nitrogen. The triggering of all potential interlocks was therefore monitored through a dummy variable across a set of new batches, and this enabled the identification of one interlock that did not work properly. This specific interlock was therefore reconfigured, and a new campaign of batches was initiated to validate the finding and assess its impact on the distribution of the batch length across the campaign.

6. Validation

After reconfiguration of the reactor safety interlock system, 8 months of operating data (corresponding to 635 new batches) were collected from the data historian. Less than 2% of the batches of the validation campaign were affected by a vacuum loss event, thus confirming that the interlock system reconfiguration was effective. In terms of distribution of the batch lengths across the validation campaign, Figure 5 shows that a unimodal distribution was obtained, with a peak value of 58 min. The distribution is also narrower than the one in Figure 2, ranging between 49 and 120 min. Therefore, reconfiguration of the safety interlock system resulted in a 29% reduction in average batch length. Overall, this amounted to an 8% reduction of the overall cycle time, with a

related saving in the energy expenses. Furthermore, it was estimated that an abnormal batch requires 25% more nitrogen than a regular batch. Considering that the abnormal batches were 40% of the total number of the historical batches, and assuming that only 2% of the batches of new production campaigns remain abnormal, reconfiguration of the safety interlock system also resulted in a 11% saving on nitrogen expenses.

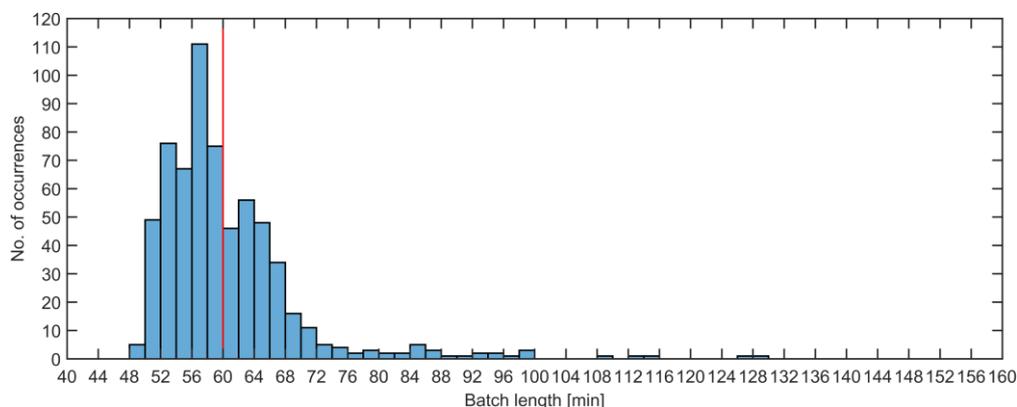


Figure 5. Distribution of the time duration of step 3 in reactor R410 after reconfiguration of the safety interlock system. The vertical red line is the mean of the distribution.

7. Conclusions

Analytics on historical data for a semi-batch industrial process, coupled to engineering understanding on the process itself, allowed us to uncover the existence of an abnormal behavior in 40% of the batches in historical manufacturing campaigns. The abnormal batches did not terminate unsuccessfully, but simply took longer to complete than the other batches. Since this event did not have an impact on the final product quality, it went almost unnoticed, and therefore did not trigger any specific action by the plant personnel. However, the longer duration increased the energy expenditure (hence, the operating expenses) per unit of product manufactured. Data analytics was central to find that the cause of the abnormal behavior was the anomalous intervention of one interlock in the reactor safety system, which (under particular conditions) caused vacuum in the reactor to be broken by nitrogen blanketing; this required subsequent vacuum reinstatement to recover the process conditions and to end the batch successfully. Reconfiguration of the safety interlock system allowed us to shorten the average batch length by 29%, and the overall process cycle time by 8%. Furthermore, an 11% reduction on the nitrogen consumption was obtained.

References

- Bieler P. S., Fischer U., Hungerbühler K., 2004, Modeling the Energy Consumption of Chemical Batch Plants: Bottom-up Approach, *Industrial and Engineering Chemistry Research*, 43, 7785–7795.
- Bro R., Smilde A. K., 2014, Principal Component Analysis, *Analytical Methods*, 6, 2812–2831.
- Brown, J. D., 2009, Choosing the Right Number of Components of Factors in PCA and EFA, *JALT Testing & Evaluation SIG Newsletter*, 13, 19–23.
- Camacho J., Picó J., Ferrer A., 2008, Bilinear Modelling of Batch Processes. Part I: Theoretical Discussion, *Journal of Chemometrics*, 22, 299–308.
- García-Muñoz S., Kourti T., MacGregor J. F., Mateos A. G., Murphy G., 2003, Troubleshooting of an Industrial Batch Process Using Multivariate Methods, *Industrial and Engineering Chemistry Research*, 42, 3592–3601.
- Jolliffe I. T., Cadima J., 2016, Principal Component Analysis: A Review and Recent Developments, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374.
- Nomikos P., MacGregor J. F., 1994, Monitoring Batch Processes Using Multiway Principal Component Analysis, *AIChE Journal*, 40, 1361–1375.
- Nomikos P., MacGregor J. F., 1995, Multivariate SPC Charts for Monitoring Batch Processes, *Technometrics*, 37, 41–59.
- Undey C., Cinar A., 2002, Statistical Monitoring of Multistage, Multiphase Batch Processes, *IEEE Control Systems Magazine*, October, 40–52.