Benefits Achieved Using Online Analytics in a Batch Manufacturing Facility

Robert Wojewodka
Technology Manager, Process Improvement Team Leader
The Lubrizol Corporation
29400 Lakeland Blvd.
Wickliffe, OH 44092
USA
(440) 347-5050
Robert.Wojewodka@Lubrizol.com

Dawn Marruchella
DeltaV Batch Product Marketing Manager
Emerson Process Management
12301 Research Blvd., Building III
Austin, TX 78759
USA
(512) 394-6123
Dawn.Marruchella@Emerson.com

KEY WORDS
Batch, Online, Analytics, Multivariate, Statistics, Process monitoring, Operational Benefits

ABSTRACT
Batch operations present manufacturers with a unique setting where operators must work in a highly complex, highly correlated and dynamic environment each day. They must also manage a large amount of data and information on a running unit – all of this making it easy for batches to end up with undesirable processing events and/or less than desirable end of batch quality. Lubrizol wanted to improve their operations by providing their operators with the ability to detect upset conditions before they have a negative impact on their batches. In order to do so, they are collaborating to develop and deploy the use of online data analytics, based on multivariate analysis, initially at their facility in Rouen, France.
Online analytics is a technology which has proven to be a challenge in batch operations for many reasons including: process holdups, access to lab data, feedstock variations, unsteady operations, data organization and concurrent batches. The benefits they have achieved were immediate and have been substantial. The analytics tools have gained the support of all company levels from management to operations staff. In addition to explaining the steps taken by the team to implement the solution, this paper will explain the modeling techniques used, the incorporation of process stages, and the benefits realized.

**PAPER**

*Introduction*

Batch manufacturing, by its very nature, presents many challenges. The goal is to always safely produce a maximum yield batch that is within product quality specifications in the shortest amount of time and with the minimum amount of waste. Not many people would argue with this goal. Not meeting it almost always equates to increased costs. This is something that everyone in the company understands – from operators to plant managers, from clerks to chief executive officers, from the sales force to the shareholders. Unfortunately, meeting that goal is not always as easy as anyone would like it to be.

The operators and engineers that work in a batch manufacturing facility are working in a very dynamic environment that is highly complex and highly correlated. In a batch facility running at full capacity, there are often multiple batches running, one in each unit. So, a problem occurring in one unit may not only wreak havoc on the batch in that unit, but could also negatively impact batches in units upstream or downstream of it. So many things are happening at once that any help with advanced warning of a potential problem or pending event is very valuable.

In addition to the dynamic environment in which they work, the operators and engineers must also manage a large amount of data and information for each unit with a batch. Data can and does come from many places including Laboratory Information Management Systems (LIMS), process control systems, enterprise resource planning systems, operations management systems, batch sheets, quality data, raw material data, and the list goes on. Even with a unit that is automated, there is only so much that can be monitored and managed at one time! Any help with automatic monitoring across many variables is valuable to the operators and engineers.

There has long been a gap between quality and control, particularly in batch manufacturing. The Process Analytical Technology (PAT) Guidelines issued by the FDA emphasize the use of multivariate analytics as a means of reducing cost and improving product quality. PAT can be defined as a mechanism to design, analyze, and control a manufacturing process through the measurement of Critical Process Parameters (CPP) which affect Critical Quality Attributes (CQA). Although this is a newer concept to many of the pharmaceutical manufacturing companies, this has been recognized as a need in many other industries and over the years has been systematically incorporated into organizations’ manufacturing excellence programs.

Quality information, fault detection and abnormal situation knowledge can be brought together to provide operations personnel with online decision support in the form of product quality predictions and early process fault detection.
The Lubrizol Corporation has expertise and a long-standing use of multivariate statistical data analysis in support of off-line process characterization and process improvement activities. This has been coupled with MIS/IT expertise in the company dealing with ERP, automation and enterprise historian systems along with the integration of these systems. Emerson Process Management had also established a research project at the University of Texas, Austin in September 2005 to investigate advanced process analytics. Both companies viewed embedded online analytics as a way to bridge the aforementioned gap between quality and control, and decided to work together in an effort to make embedded online analytics for batch processes a reality.

**Objectives**

In order to effectively address the embedded online analytics problem, a set of batch analytic tools was developed that would help to improve plant operations. It was important that the model building tools be easy to use. That is, the process engineer, chemist, or process analyst should be able to easily define, update and evaluate the models. Perhaps more important was that the online evaluation tools be intuitive enough for operators to use and understand with the following goals in mind:

- At any point in time as the batch is being made, indicate to the operator if that batch is in statistical control across all process variables as it relates to the variation characterized from past successful batches. Should it start to deviate, this may be considered an alarm.
- Provide a mechanism to drill down through the relevant process variables related to this deviation to allow operators to identify the reason and take action.

The tools and technologies developed were used in a field trial in one of Lubrizol’s manufacturing facilities located in Rouen, France. The primary objectives of the field trial included:

- demonstrating the online prediction of product quality and
- evaluating different means of online process fault detection and identification.

The results of the field trial were analyzed to see what was learned such that the tools and technologies could be updated and improved. The results and benefits of the technology have been documented, in order to determine if this is something that could continue to be used throughout the company and ultimately by others as well.

**Challenges**

As stated earlier, batch manufacturing presents many challenges. So, one can only expect that the application of analytics to batch manufacturing also presents many challenges. Some of the major issues that need to be addressed are discussed below.

Process holdups are a common occurrence in batch manufacturing. These holdups can be normal parts of the process or due to upset conditions. For example, as part of the process, an operator may be prompted to perform a specific activity. While performing that activity, the batch sits and waits. Any time a human is involved in the process, there will not only be holdups, but also variations in the amount of time it takes the operator to perform the activity before the batch can progress. An example of an unplanned holdup is a batch being put in the Held state for any number of legitimate reasons. The length of time in which it remains in this state will vary. Tools must account for operator and event initiated processing halts and restarts. Dealing with the alignment of data from batch to batch is discussed in the **Data Analytics** section of this paper.
Access to laboratory data is important not only for use in calculations by the operator or automation system, but also for use in analysis of the process. Lab analyses results are typically in a LIMS system or in ERP systems and may not be easily accessible for use as part of the analytics tools or in calculations. Furthermore it is not always practical to have an inline analyzer for measurements. For example, lab results indicating the concentration of a sample taken from the process are needed in order to accurately make adjustments to the process or to amounts of other raw materials needed for the manufacture of the batch. *Lab results must be available to both the off-line and the online analytic toolsets.*

Variations in feedstock are inevitable, especially when a supplier is switched (unbeknownst to the operations personnel) and often times will require adjustments to the process. For example, when a new shipment of material is added to a storage tank that supplies raw materials to a batch process, it is critical that the impact of the addition of that material and its relevant parameters is known. The properties of the contents of the storage tank have been altered and could impact future batches in which those contents are used. This data may not be readily available to the analytics tools as it could be something kept by purchasing. *The properties associated with each material shipment should be available for use in online analytic tools.*

Varying operating conditions will occur as a batch progresses through its different stages and is moved from unit to unit. In order to be accurate and effective, the models and tools must be aware of the different equipment and process steps for a particular batch. *The analytic model must account for the batch being broken into multiple operations that span multiple units (equipment).*

Concurrent batches are common in almost any manufacturing facility in order to maximize equipment utilization and product throughput. As soon as one batch is completed in the first stage of a unit, the unit is prepared to initiate the next batch. This means that there will no doubt be many batches active that the operator will need to monitor, making it necessary for the tools to be able to differentiate the various batches. *The data collection and analysis toolset and online operation must take into account concurrent batches.*

Assembly and organization of the data is a requirement that must be met in order to accurately perform a process analysis and to use it in an online fashion. *Efficient tools are needed to access, correctly sequence, and organize a data set to analyze the process and to move the results of that analysis online.*

**Data Analytics**

Solutions exist to address each of the individual challenges listed above. Some of these are available as commercial products while others are custom solutions. Because of this, finding an analytics package that takes the above into account proves to be a challenge.

The factors impacting a batch process are depicted in Figure 1. With each process, there are one or more inputs that can all have variability. Things such as supplier, raw materials, time in storage, actual quantities used and potency of those materials are typical contributors to input variability. While the process is running, conditions can also vary. The equipment used, cycle time, operator on shift and environmental factors are all things that will impact the operating conditions of the process. Control and output variables are also used and can have variability. The key to manufacturing a good batch, i.e., one that meets product specifications and maximizes yield, is to be able to manipulate the correct variables based on both measured and unmeasured values.
Some companies determine the quality of a batch by comparing that batch to a previously run batch that has been marked as the “Golden Batch”, or a batch that has met quality standards and is considered to be ideal for a particular process. While seemingly acceptable, this approach can be extremely problematic in that when only one batch is used as the only acceptable batch, it doesn’t take into account the interactions between inputs and process variables. This is actually taking a univariate approach to what is actually a multivariate problem. Comparing a batch to a multivariate-based model, or a “Golden Profile”, for a particular process that is based on a compilation of many “Golden Batches” allows for more accurate analysis of that batch, and depending on the techniques used, fault detection and end of batch quality prediction. In addition, process insight is obtained which is not obtained at all from what many other companies promote as a “Golden Batch” analysis.

Figure 2 shows a simple example of how an operator would not be able to distinguish a problem for normal operations if he were just looking at a single variable in the chart. All of the points in the figure appear between the accepted upper and lower control limits from a univariate perspective. However when looking at this from a multivariate perspective, it is obvious that there are three points that deviate from the modeled relationship. Imagine what this would look like in a typical process with many variables.
By applying statistics to this model, these abnormal conditions could be easily identified on a single trend, even for scenarios with many variables. An example is shown in Figure 3.

In addition to the challenges mentioned previously, data analytics must also be able to accomplish the following objectives in order to be successful:

- Take a very large number of input and process variables associated with a batch process and characterize “acceptable variation” and process relationships.
- Identify how these variables relate to each other and to end-of-batch product quality characteristics.
- Be used to identify typical and atypical process relationships as current and future batches are running.
- Be used to predict end-of-batch quality characteristics at any point in time as a batch is evolving.
- Provide additional and focused information to operations personnel to improve batch processing during production.

A brief discussion of the primary multivariate and statistical methods used in the analytical models follows.
Principal Components Analysis (PCA) is a technique used to provide a concise overview of a data set. PCA is powerful for pattern recognition in data such as identification of outliers, trends, groups and relationships. It is through the use of PCA that it is possible to detect abnormal operations that result from both measured and unmeasured faults.

Hotelling’s T2 statistic is used to quantify measured disturbances where the Squared Prediction Error (SPE) or Q statistic is used for quantifying unmeasured disturbances. A fault is then determined by comparing both of these calculated statistics to an upper limit. An abnormal condition is indicated if the value exceeds the limit (as is shown in Figure 3).

Projections to Latent Structures (PLS) is used to establish relationships between the input and output variables present in a batch process and is also used to develop the predictive model for the process.

PLS with Discriminant Analysis (PLS-DA) is powerful for classification and is used to create the predictive model for the process where the objective is to accurately classify categorical attributes of a batch.

Before the process can be modeled, however, the data must be aligned. To help visualize this, the chart on the left in Figure 4 is an overlay of the unaligned data for a parameter trend for many batches. Each line represents the value of that parameter throughout a different batch. Without aligning this data, building a model or attempting to determine whether or not a batch should be used in model generation would be difficult to do. There are several approaches to data alignment; one approach is to just chop off the data at a certain point in time and another is to “squeeze or expand” the data in an accordion-like fashion. It was felt that neither of these approaches was technically the best way to handle data alignment and so a technology that was originally developed for use in speech recognition technology known as Dynamic Time Warping (DTW) was used to align batch lengths. The chart on the right in Figure 4 shows the same data as that on the left with DTW applied to align the data.

![Figure 4: Data Alignment using Dynamic Time Warping (DTW)](image)

Field Trial

Several steps were taken in preparation for the actual field trial and will be discussed briefly here. A multi-discipline team which included plant operations was formed and involved throughout the process. Team input was captured using an input-process-output data matrix to help with the analysis.

As mentioned earlier, it was critical that lab data as well as truck shipment data be integrated and available for use by the analytics package. Feed tank properties needed to be calculated automatically with each new addition of raw materials from truck shipments.
A survey of all process instrumentation and loops was performed to ensure that all of the instrumentation used was without problems and that loops were tuned to achieve optimum performance. The idea here is that in order to apply advanced control to the process, one needs to make sure that the basic equipment is properly functioning first. To not do so could lead to wasted time and effort in the long run.

Before going online with the trial, operator training was held so that the operations personnel were comfortable enough with the new application to use it in the field.

The actual field trial was run on two batch manufacturing processes. Each process made different products. In addition, the output of process 1 was also an input into process 2, as outlined in Figure 5 below. A total of eighteen input variables, thirty-eight process variables and four output variables were used for purposes of the process modeling and online analytics. Finally, a total of one hundred and seventy-two historical batches were used for purposes of analysis and model development across these two processes. In reality many more process variables would be included. For the purposes of this trial the scope was kept small.

Figure 5: Actual Field Trial Processes

ISA S88.01 defines a stage as “...a part of a process that usually operates independently from other process stages and that usually results in a planned sequence of chemical or physical changes in the material being processed”. Consistent with that definition, the analytic models are defined for each stage thus allowing for the inputs and outputs used in analysis to be different for each stage. A separate model exists for each stage – defined uniquely by product, equipment and process operations.

The tools used for the development of the models offline required selection of historical batches to be used, along with selection of appropriate variables from historical data. Refer to Figure 6. The tools also provided for the ability to compare the results of the model against historical data in order to determine the accuracy of the model or if a particular batch was an outlier and should be removed from the selection of those used in developing the model.
Once satisfied with the models, the online analytics tools were put to use in a web-based interface. It was important that a web-based interface be used as some of the process specialists who would be working with the analytics were located at other locations and sites throughout the world. With this approach, users only need to have Internet Explorer, a connection, and privileges to the network upon which the analytics was running.

Figure 7 shows the main display monitored by operators. For each process, active batches are displayed along with an indication of whether or not there is or has been a process fault detected. When the operator sees an indication that a fault has been detected, he simply selects the batch and is taken to the display shown in Figure 8. On this display, the operator is shown the statistical charts for the selected batch. Anytime the statistical values for the batch exceeded the upper limit (anything greater than a value of one is considered to be statistically a fault), the trend for the indicators appears outside of the gray zone. The operator can select anywhere on the line and see on the right side of the display the top five contributing variables for this point in time. All statistical terminology is removed from this display and only terms familiar to the operations personnel are used. This way, complex statistical calculations and terminology may be kept “behind the scenes”, resulting in a more intuitive interface for the operators and engineers.
The operator may also view the contribution of all of the variables associated with a fault for any point in time by simply going to the Contribution tab shown in Figure 9.
Finally, the operator may select any one of the variables and will be taken to the individual trend for the selected variable in order to further analyze the situation. This is shown in Figure 10. The operator is able to easily see the trend of the variable in question for the selected batch overlaid against not a single golden batch, but against the model that was developed for the particular process, along with the upper and lower limits for the variable. In the example shown, it is clear that the variable being trended is well below the output from the model and the acceptable variation, so that the operator can address the situation as needed.

Finally, the operator is able to view on a separate trend the predicted end of batch quality for a selected batch, with confidence limits.
**Benefits Realized**

The benefits that have been realized at the Rouen, France facility have been numerous and ongoing. They started immediately. Some of the benefits are highlighted below.

- As part of preparation for the field trial operator training was done. Due to language barriers – not everyone was bilingual in English and French – a train the trainer class was done. During the class a fault in the actual process was detected using the online analytics. While this halted the class temporarily, a previously undetected problem was identified with the mass flow meter for a key component that is charged into the batch. This fault was *going unnoticed with “traditional” monitoring systems*. This helped to highlight the benefits of this technology and certainly got the attention of operations management!
- Once up and running, the field trial showed through several batches an increasing deviation on the key component density measurement. This phenomenon was linked to the start of plugging which was quickly solved by applying steam *without time cycle impact*.
- The on-line tool indicated a problem with the cooling system of the reactor. It detected that another key component charge was being introduced too slowly and that the reactor temperature was running a little bit higher. The problem was solved on the cooler. *Using a univariate approach would not have shown this.*
- A process fault detection led to the identification of a regular issue on the unit’s reactor heating control loop. This led to actions to *revisit the loop tuning on key process control parameters*.
- Another problem was detected on the hot oil heating system. The problem identified was with an upstream boiler that was negatively impacting operations. A quote from the operations personnel follows:

  “…thanks to the Beta. An equipment failure was discovered in advance and (the plant) *avoided loosing 5 hours per batch* for the batch in process and also for the following batches before discovering the problem with the traditional manner. Probably some days would have be(en) necessary to discover that type of mechanical problem without the Beta. (Boiler combustion air controller located in a bad accessible zone and thermal oil leakage). …(we would have) discovered this latter with the periodic update of the indicators of efficiency, but we saved time earlier thanks to the beta. Earlier is better than too late!”

**Status and Next Steps**

At the time of this writing, the primary field trial at the Rouen facility has been successfully completed. Work is underway to update and redeploy the online models for the two manufacturing units in the plant. These models will continue to be used until a commercialized product is available.

Further work is also being done in the form of a design of experiments being done to evaluate this functionality in order to further explore any nuances of the methodology that could be incorporated into the commercialization of the modules in order to further improve their capabilities.

Finally, another field trial is underway for a batch process in a different process and different industry in order to determine what additional modifications, if any, should be made. Work toward commercialization of the product has been initiated.
Summary
Statistical data analytics allows the engineers and operators to easily see when there is a potential problem with a batch and to further determine if that problem will have an impact on that batch’s quality. The more that they use tools such as those discussed in this paper, the more they will begin to understand the interactions between variables in their process. Process chemists and engineers will have more information available to them to make additional process improvements.

As the use of online data analytics becomes more prevalent, a greater understanding of the real components of a batch process will result not only in an increase in the consistency of quality products, but also in an increased throughput through better yields and decreased cycle time and outages. In short, it will result in more good batches at a lower cost to manufacturers. If the results of the Rouen field trial hold true for others, manufacturers will derive additional insight and benefits in their manufacturing by use of online analytics in their facilities.

References and Additional Information