CONTINUED PROCESS VERIFICATION IN STAGES 1–3:
Multivariate Data Modeling Using Design Space and Monte Carlo

By Zuwei Jin

Continued process verification (CPV) as defined in the US FDA process validation guideline [1] helps bring quality management and compliance in the pharmaceutical industry to the next level, but it has been challenging to implement in practice. This article describes an approach for implementing CPV through the core concept of design space based on online multivariate data analysis (MVDA) and Monte Carlo random simulation.

T he approach can use virtually any kind of data source to build the design space, including first-principle dynamical models, design of experiment (DOE) models, clinical trial batches during process performance qualification (PPQ), and historical batches in a production historian. This approach can provide a smooth transition from the research and development (R&D) first-principle model to the permanent CPV program for commercial production throughout the drug development cycle.

The FDA recommends a three-stage approach to process validation. As the US pharmaceutical industry regulator, the FDA has been driving science- and risk-based approaches for almost a decade through documents such as the process validation guidance published in 2011 [1]. The guidance fundamentally affects process development, engineering practice, and commercial production for drug substances and drug products. A new commercial manufacturing process should go through stage 1, process design; stage 2, process qualification; and stage 3, CPV. Whereas stage 2 retains most procedural elements from traditional qualification and validation (such as installation qualification, operational qualification, performance qualification, and process validation), stages 1 and 3 involve many science-, risk-, and statistics-based approaches,

such as risk assessment, DOE, statistical process control (SPC), and processing capability (Cpk) evaluation. The FDA now recommends stage 3 for all commercial processes because it provides the ultimate evidence that a process is running under a state of control.

The FDA's process validation guidance [1] also emphasizes statistics. Sponsors are encouraged to identify critical process parameters (CPP) and critical quality attributes (CQA) through risk assessment according to the quality target product profile (QTPP) and evaluate them using statistical approaches such as DOE early in stage 1. The historical standard of three consecutive batches may no longer be sufficient for chemistry, manufacturing, and controls (CMC) submission. The FDA is now considering the entire drug development approach, how much product and process understanding the drug manufacturer has demonstrated, and statistical evidence that the process is running as designed and in a state of control. Including CPV as part of CMC submission is therefore highly recommended.

CPV is now included as part of annual product review. SPC charts and Cpk analysis are the most common tools used in current CPV programs. They are univariate methods and are usually used after batch completion. SPC charts for quality attributes are usually used to evaluate processing capability and may also be used to assess control capability for specific process parameters. Cpk analysis evaluates control system capability by monitoring variation of the process parameters. From a statistics standpoint, far more than three batches will usually be needed to sufficiently analyze process capability of both process parameters and quality attributes. Therefore, a CPV program using univariate tools would usually not be established until phase 3 commercial production.

Because of its complete statistical analysis capability, MVDA is sometimes used to better understand the correlations between the CQAs and CPPs in place of SPC and Cpk. Such analysis would, however, still have to be done after batch completion, and extensive modeling and computation would usually be involved. The

obvious advantage of MVDA is its ability to count the interactions between multiple process parameters that a univariate approach would fail to detect.

In following sections, an online MVDA–based approach is introduced to allow CPV to be planned as early as stage 1; this approach can make CPV an integral part of process validation throughout stages 1, 2, and 3. It provides not only the ability to statistically evaluate the correlation between the CQAs and CPPs but also the ability to detect process fault and predict CQAs in real time.

ONLINE MVDA

Online MVDA is a relatively new data analytics technology that can be used for CPV. Online MVDA is based on the traditional MVDA methodology and a batch simulation scanning concept. Before we discuss the challenges of implementing CPV for the entire drug development cycle, let us briefly review how online MVDA can be achieved.

Batch Simulation Using MVDA

MVDA is a powerful improvement to the current univariate approach in CPV and is recommended by many industrial experts [2]. Primary component analysis (PCA) and projected latent structure (PLS) in MVDA are highly recommended in data analysis. With modern computing power and iterative PCA/PLS algorithms, both enumerative and analytical statistics can now be achieved efficiently online in real time.

How can MVDA be used for a batch process, which is dynamic and can't be directly analyzed with MVDA?

MVDA is complicated with its high number of dimensions, but batch simulation is complicated in its own way because of its dynamic nature. Batch simulation using MVDA requires an additional concept or data structure. Such a data structure would allow for sufficient description of any batch process in a time-evolving fashion. In practice, MVDA modeling for batch would involve scanning a batch process into many frames over the batch duration (Figure 1). These frames are called statistical scans. The statistical scan averages the process conditions between two nearby frames. A batch process can be represented by many statistical scans that are sometimes separated into several groups called stages [3].

Within the stages, statistical scans may need to be appropriately aligned from batch to batch. One of the challenges of batch simulation is that batches may be of different durations and pauses/holds may occur during them; therefore, the alignment of statistical scans can't be based on time only. A well-accepted approach to addressing this challenge is an algorithm called dynamic time warping (DTW). DTW is an optimization that looks not only at the time but also the physical characteristics—such as process parameters—of a batch process to determine the alignment of the statistical scans [4].

Batch MVDA models contain hundreds of MVDA models that are lined up at different time frames in the scanning structure. The amount of modeling in a batch simulation is a hundred times

Figure 1: Batch simulation using MVDA.
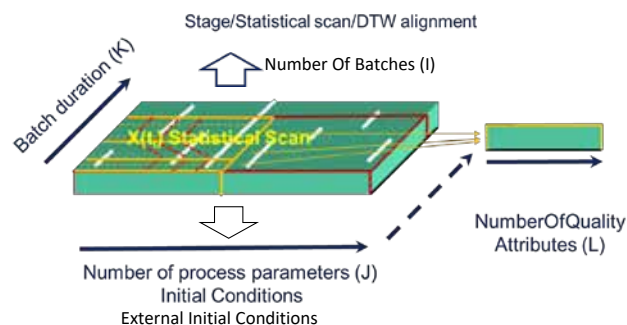
greater than what would usually be modeled in an R&D project. Batch simulation with MVDA is not be feasible without an automated platform.

Automating MVDA

MVDA can be fully automated online to provide real-time analysis thanks to the modern computing power of servers, the standardization of plant models as defined in ANSI/International Society of Automation (ISA) Standard S88 [5], and digitization of almost all process parameters and initial conditions. Algorithms such as NIPALS, which is an iterative decomposition algorithm for matrix data, make PCA/PLS an effective online method for analyzing large amounts of data and building MVDA models.

Another important foundational piece of online MVDA is the algorithm to align statistical scans from batches of different durations. One of the most popular approaches is to use DTW, which takes a certain number of adjacent statistical scans into consideration during alignment. MVDA models can therefore be generated automatically from historical data in manner similar to the way Google performs an internet search.

Online MVDA is an integrated part of many common process control systems, such as distributed control system (DCS). An online MVDA platform positioned on top of historians includes a model builder, an analytics server, and a monitoring server (Figure 2). The model builder builds MVDA models from historical data. The monitoring server uses appropriate MVDA models to evaluate the actual performance of a real-time process and predict batch quality. The online MVDA platform sits on top of a process control system at level 2.5 or 3 in the ISA S95 hierarchy.

Online MVDA incorporates built-in model-building tools such as PCA and PLS as well as a real-time monitoring server based on the ISA S88 structure to provide the capability to detect faults and predict quality in real time. The fault detection and quality prediction windows are essentially the automated form of CPV, continuously measuring real-time process against the design space.
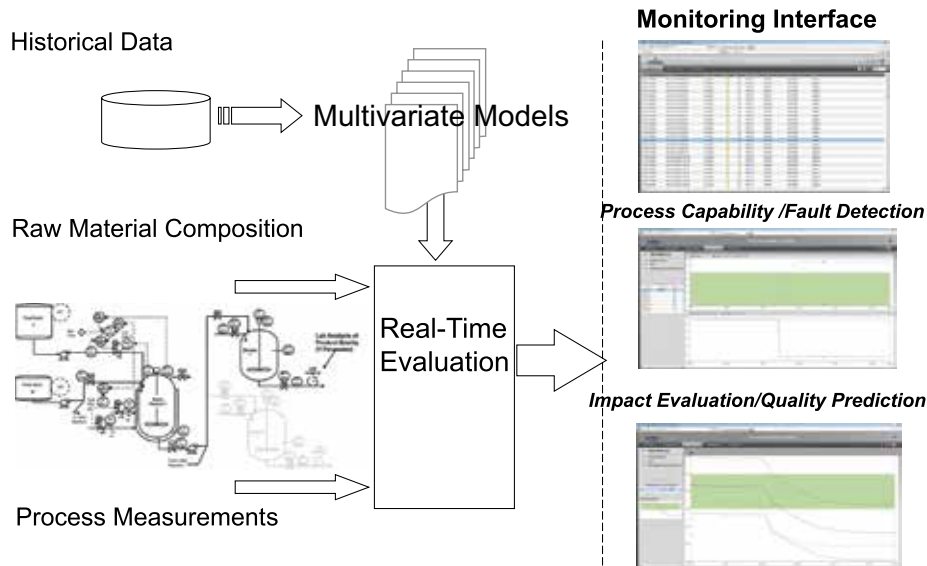
Compared to traditional biplots in PCA, normalized T^2 and Q (between 0 and 1) are a more statistically robust method to identify irregularity and ensure a process is running under a state of control. In addition to providing enumerative statistics such as those used in CPV, online MVDA provides real-time monitoring functions such as process fault detection and batch final quality prediction.

In Figure 2, the middle window on the right illustrates how the monitoring service performs statistical tests on process parameters, specifically on T^2 and Q based on PCA. Statistically improbable behavior will be called out. Specific contributions from different process parameters can be further investigated from the same window by clicking the process parameter contribution on the left. The bottom window on the right shows the quality prediction (middle line) for the real-time batch. The top and bottom lines represent the upper and lower limits, respectively, of the prediction, all at 95% confidence level. Operators may be trained with standard operating procedures to use the fault detection and quality prediction tools to intervene in the process when required.

IMPLEMENTATION CHALLENGES

Figure 3 shows the landscape of process simulation and current CPV practice in the pharmaceutical industry. Implementation of a golden (i.e., ideal) batch profile is desirable to comply with industry regulations and achieve operational excellence. The challenge, however, has been the disconnect between the dynamic nature of the R&D first-principle model, the statistical nature of DOE design space, and the actual discrete data from real-time processes. Most of the current control platforms do not allow the manufacturer to set a golden batch profile as the background for batch operations, particularly at process start-up. In the case where a golden batch profile with standard deviation can be provided by the historian, such monitoring is usually univariate and not statistically meaningful enough to detect fault or predict quality.

Design space is the common thread that connects all these pieces together, but there has been no common form for representing it from stage to stage. The pharmaceutical industry generally lacks platforms that can fully integrate R&D, technology transfer, clinical manufacturing, commercial operations, regulatory compliance, and manufacturing intelligence.

The online MVDA model has been successful with production monitoring, but implementing it for CPV throughout the phases is challenging. It takes far more than three actual batches to build an MVDA model through PCA/PLS; therefore, MVDA historically could not be used in early stages because there were not enough batches. Furthermore, the R&D DOE design space and first-principle mechanistic models early in process development can't be directly used for building the MVDA model because such equations can't be included in an MVDA platform. Most MVDA online platforms do not take in loose data or equations for model-building. Thus, CPV in current practice is mostly SPC charting and quality parameter trending at clinical phase 3 and commercial phase 4, and is completely disconnected from earlier stages.

IMPLEMENTING DESIGN SPACE USING MONTE CARLO SIMULATION

An application of Monte Carlo simulation and open data sources for MVDA model-building can bridge the gaps discussed previously. It allows online MVDA to be a solution throughout the entire

Figure 3: Disconnects in current technology transfer practice.
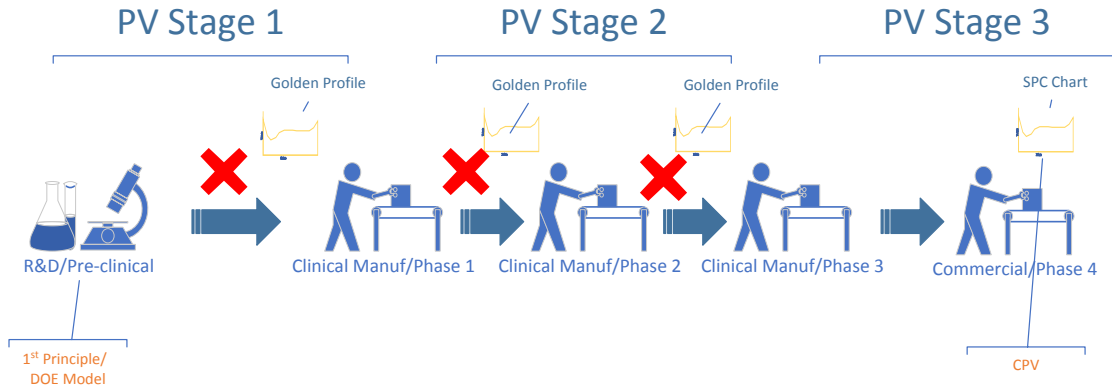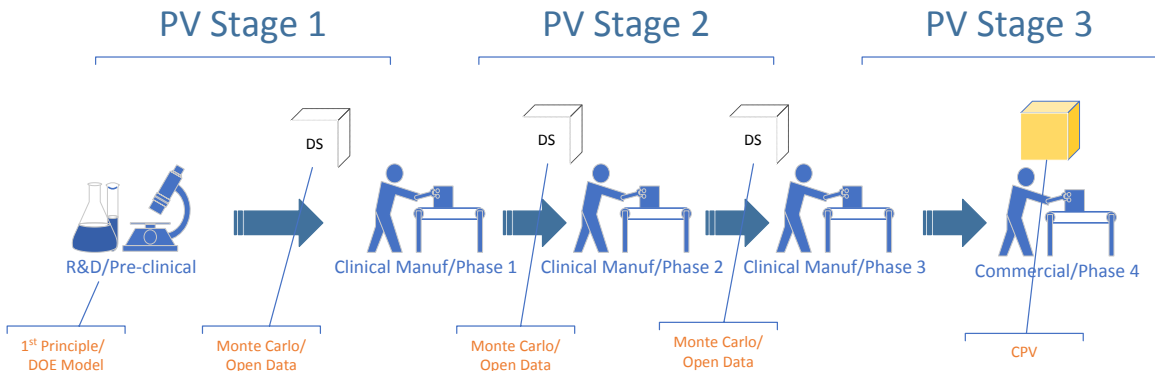

Figure 4: Implementing design space through stages 1, 2, and 3 for CPV.

drug development cycle from process development to commercial production, and that solution is applicable to the entire plant [6].

Although batch history data are highly structured in the historian, they can also be randomly simulated as new batches within the design space. With a Monte Carlo simulation framework, the design space and a reference batch can be used to construct as many batches as desired. The parameters required in the Monte Carlo simulation are the approved parameter range, the parameter control capability, and a reference batch.

The simulated batches are initially used to build the MVDA model and gradually replaced with real batches from commercial production to rebuild the model. In this way, the MVDA model improves over time. CPV can be achieved when all simulated batches are replaced with real batches.

Allowing the use of open data sources means that the model definition can be defined not only through the batch historian but also through ISA S88 hierarchy exported from a DCS or other control systems, separate batch data files, and batch event files. This means third-party historians and even data from a first-principle mechanistic model can all be used for MVDA model-building on such platforms.

The way to implement a first-principle or DOE model for production is not to take in the equations directly but to take in a sufficient amount of data from the DOE or first-principle equations for the MVDA model to learn. When a process is transferred from one phase to the next, the design space may have to be transferred using Monte Carlo simulation with an open data source because the process will be built anew at a different scale (see Figure 4).

With the Monte Carlo capability and the open interface in building batch context, any golden batch profile or DOE design space—regardless of where it came from—can be implemented as an MVDA model for process monitoring. Figure 5 shows how an MVDA model may be constructed and deployed using the Monte Carlo method and an open data source.
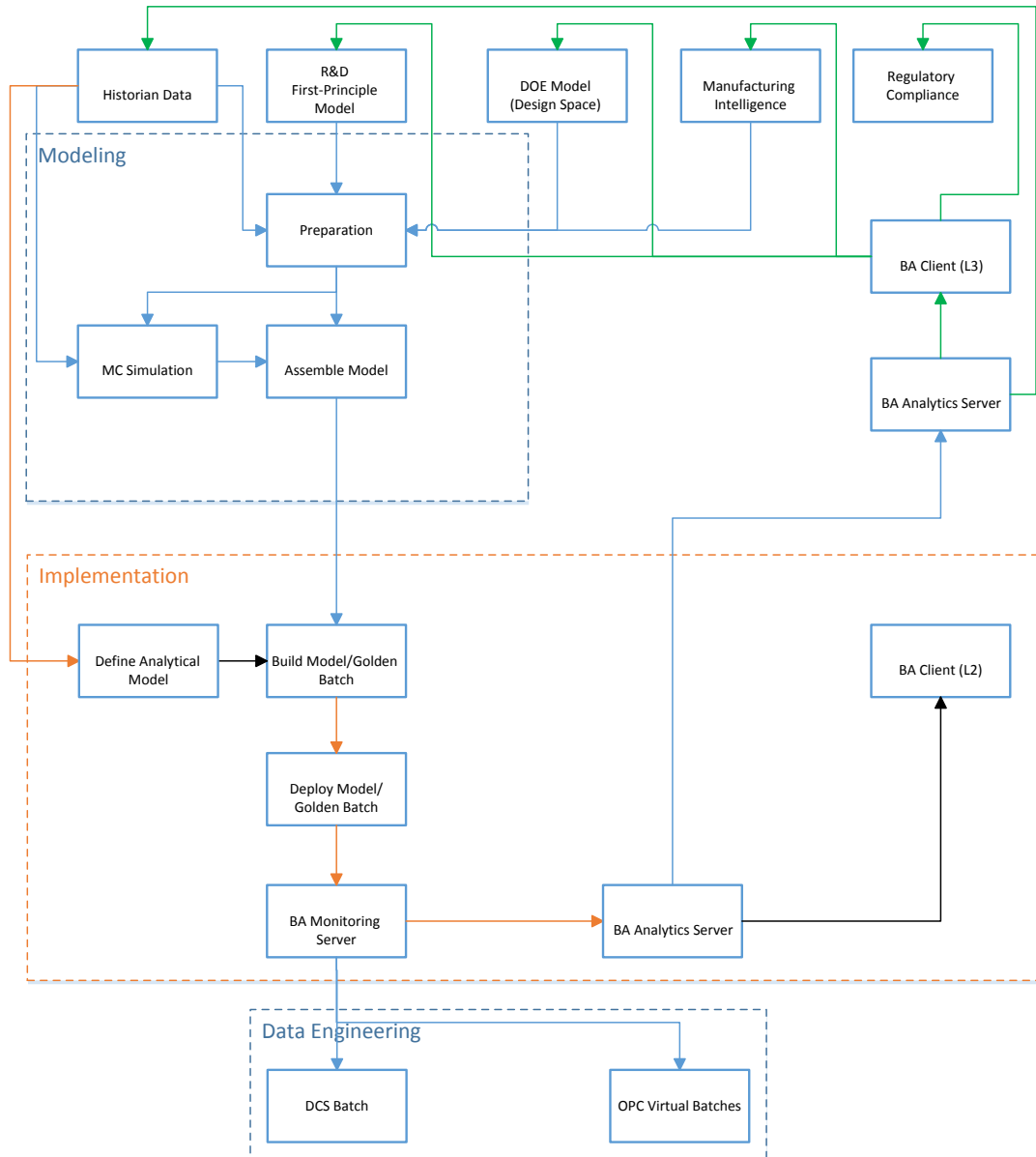
The MVDA model in this case can be built from the historian, the DOE models, or first-principle models. If simulated batches are needed for model-building, Monte Carlo simulation must be performed to generate such batches before a complete MVDA model can be assembled. Process parameters and quality attributes must then be defined for the batch, and PCA/PLS are run to generate the specific characteristics of the MVDA model, such as eigenvalues, loading, and scores. The MVDA model can then be deployed to the specific control platform for real-time monitoring. This effort must be coordinated with plant automation. As shown in Figure 5, the foundation for MVDA model deployment and model-building is data connectivity within the plant.

IMPLEMENTING ONLINE MVDA FOR THE ENTIRE PLANT

From a business standpoint, implementation of online MVDA involves two challenges: First, the manufacturer wants to find an analytics solution for the entire plant, if not the entire enterprise. Second, the solution should function not only as an operation support tool but also as a manufacturing intelligence tool that supports business decisions.

Thanks to the Open Platform Communications (OPC) connectivity standard, the MVDA online platform can be applied to all control systems in a plant. Most modern control systems have open connectivity, such as OPC, which can allow the control system to have virtual batches in the MVDA online platform; those virtual batches can then be monitored just like regular batches by online MVDA solutions.

MVDA models can be built for a specific product on a specific unit. The MVDA online platform can therefore be used for multiple products on nondesignated equipment.

MVDA can be a powerful tool for both operations and business decisions. MVDA analytics can be used for decision support, manufacturing intelligence, or quality assurance for business improvement. Most MVDA online platforms have the capability to work across different domains and can be configured through network firewalls.

IMPLEMENTATION COST OF ONLINE MVDA

An MVDA online in a production environment is generally assumed to be expensive because of MVDA licensing costs and the complexity of project implementation. Although online MVDA is by far one of the most powerful analytics tools available, the threshold for adopting it is still high.

> MVDA analytics can be used for decision support, manufacturing intelligence, or quality assurance for business improvement.

Implementing MVDA throughout a drug development pipeline requires long-term collaboration among experts in operations, process engineering, automation, statistics, data sciences, and information technology (IT). For example, experts in IT must help with data connectivity, and model-building requires process-modeling and statistics knowledge.

Benefits of Membership

Network with more than 17,500 members in 90+ countries to gain industry knowledge, learn best practices, connect with peers, and advance your career with the largest not-for-proft association in the pharmaceutical industry. **ISPE.org/Join.**

A Community of Experts

Direct access to a global network of like-minded experts, through the ISPE Member Directory and our newly relaunched online community portal, Community Connection.

Affiliates and Chapters

Get involved with your local Affiliate or Chapter to meet local industry peers, volunteer your time, and mentor Young Professionals and Students.

Guidance Documents

ISPE Guidance Documents are the gold-standard in the industry. Members save 60% off nonmember prices, and enjoy FREE online access to select ISPE Good Practice Guides.

Currently excludes GAMP® and Baseline® Guides

Educational Webinars

Enjoy unlimited access to ISPE's complimentary Pharma Best Practices Webinar Series featuring leading subject matter experts covering critical, relevant topics in pharmaceutical manufacturing.
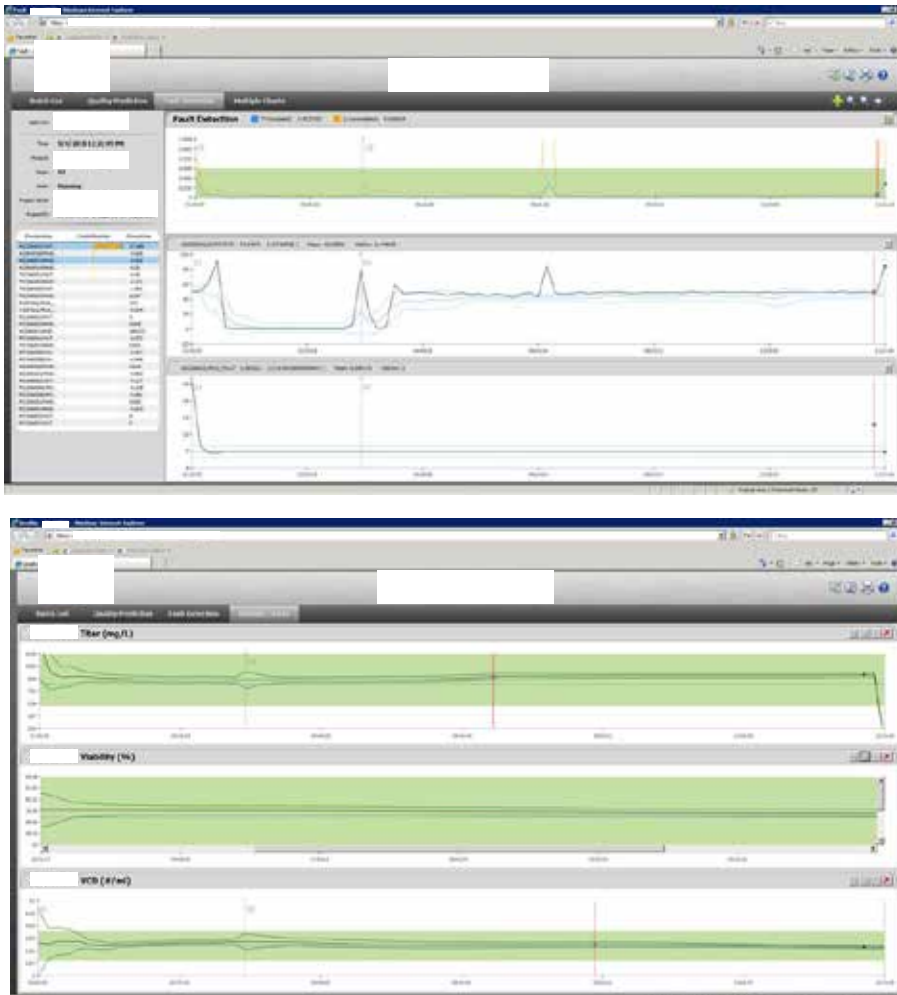
Figure 6: **An example of a golden batch profile implementation through MVDA modeling.**

Figure 7: **Real-time monitoring of the batch process using an MVDA model.**

Online MVDA licensing costs may vary significantly from vendor to vendor. Some MVDA online solutions have been highly integrated with a particular control system, which could reduce the implementation threshold for that control platform; however, integration with third-party control systems may be difficult.

CASE STUDY

Figure 6 shows a golden profile/MVDA model built from a reference batch and a design space. This ideal batch profile is an active statistical model simultaneously used for statistical evaluations of the actual running batch based on T^2 and Q, as explained previously. When the running batch is not performing normally, T^2 and Q will signal that issue by showing the values moving out of the green area.

A production process, 5K bioreactor cell culture batch, in a large US pharmaceutical manufacturer's facility was used as a proof of concept for this approach to implementing design space using Monte Carlo simulation. Simulated batches were generated using the Monte Carlo method based on reference historical data from an IP21 historian and a design space available from earlier development work.

The simulated batches were then used to build an MVDA model, which included 26 process parameters as the process input and three quality attributes as output. The model was then successfully deployed for online real-time monitoring in production. The monitoring results are shown in Figure 7.

The MVDA statistical model calculates T^2 and Q on a periodic basis to monitor real-time batch processes. Essentially, the real-time batch is being constantly measured against its design space, which is represented by the MVDA model built through Monte Carlo simulated batches.

The top windows in Figure 7 show T^2 and Q for fault detection of the process conditions. The bottom window shows the conventional univariate golden profile or historical behavior for one of the selected process parameters. T^2 and Q statistics plus the univariate golden profiles are the design space of the process, which can then be used to measure the running batch in real time.

As more batches are completed, the MVDA model is rebuilt with the new batch history data as part of the training or testing batch data set. This rebuilding process continues until all simulated batches are replaced with real batch history data.

CONCLUSION

Online MVDA is emerging as the solution for CPV throughout the drug development cycle, connecting stages 1, 2, and 3 and allowing CPV to be included early in the drug development life cycle.

Design space is the common thread that connects stages 1, 2, and 3 in process validation. MVDA modeling with Monte Carlo simulation and an open data source for model-building is one of the approaches moving the design space of the drug-making process from stage to stage and from phase to phase to achieve CPV, as the FDA recommends in its process validation guidance.

Although equations from first-principle and DOE models can't be used by an MVDA platform, first-principle models and DOE design space can be used to provide the data that the MVDA model builder uses to learn about the design space of the first-principle or DOE model.

The online MVDA solution is particularly suited for handling large amounts and complicated sets of data in plantwide applications, and it can be applied to different control platforms with OPC connectivity. Analytics by online MVDA can support not only operations but also quality assurance, compliance, and manufacturing intelligence.

Online MVDA can be used for CPV throughout the drug development cycle, connecting validation stages 1, 2, and 3. It is arguably the future direction of process monitoring in the pharmaceutical industry. ✏

References

1. US Food and Drug Administration. "Guidance for Industry. Process Validation: General Principles and Practices." 2011. https://www.fda.gov/media/71021/download

2. BioPhorum. "CPV—Continued Process Verification Case Study." 2015. https://www.biophorum.com/cpv-continued-process-verification-case-study

3. Reiss, R., W. Wijiznis, and R. Wojewodka. "Partial Least Square Confidence Interval Calculation for Industrial End-of-Batch Quality Prediction." *Chemometrics and Intelligent Laboratory Systems* 100, no. 2(2010): 75–82. doi:10.1016/j.chemolab.2009.11.003

4. Wijsznis, W. K., and T. L. Blevin. On-line Alignment of a Process Analytics Model. US Patent 20110288660, filed 21 May 2010 and issued 4 November 2014.

5. International Society of Automation. *ISA Standards.* https://www.isa.org/standards-and-publications/isa-standards

6. Jin, Z. Methods and Apparatus for Using Analytical/Statistical Modeling for Continued Process Verification (CPV). US Patent 10394973, filed 18 December 2015 and issued 27 August 2019.

Additional Resources

Abdi, H., and L. J. Williams. "Principal Component Analysis." *Wiley Interdisciplinary Reviews: Computational Statistics* 2, no. 4 (2010): 433–459.

Bengio, Y., A. Courville, and P. Vincent. "Representation Learning: A Review and New Perspectives." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, no. 8 (2013): 1798–1828. doi:10.1109/TPAMI.2013.50

Jolliffe, I. T. *Principal Component Analysis.* New York: Springer-Verlag, 2002.

About the author

Zuwei Jin is a Senior Industry Consultant with Zenith Technologies. He specializes in manufacturing execution systems (MES), data analytics, and advanced process control (APC). He previously worked for Emerson Automation Solution for six years and GE Healthcare Life Sciences for 16 years. Zuwei's technical experience includes biological production processes, cGMP hygienic equipment, process automation, data analytics, and operation management automation solutions. Zuwei has been involved in many engineering projects in the life sciences industry and is well versed with the best practices and various industry standards relevant to pharmaceuticals and automation. He has a PhD in chemical engineering from Ohio State University. Zuwei has been an ISPE member since 2013.