ENHANCING CHO PROCESS UNDERSTANDING FROM CHO MANUFACTURING PROCESS DATA

Thomas Mistretta, Amgen, Cambridge, MA
tmistret@amgen.com
Patrick Gammell, Amgen, Cambridge, MA
Cenk Undey, Amgen, Thousand Oaks, CA
Greg Naugle, Amgen, West Greenwich, RI

Key Words: Big Data, Data Lake, Predictive Models, Process Monitoring, CPV

It has been said that the biggest data can come from the smallest packages and this is certainly true in the case of CHO cells, CHO-based commercial bioprocesses can generate up to 500,000,000 data points per batch. When aggregated, organized, and analyzed, these data represent a significant, but generally underutilized, opportunity to advance product understanding and process control opportunities.

Traditional approaches to CHO process analytics have typically leveraged discrete in-process control and analytical release test data for batch over batch trending to ensure that processes remain in a state of statistical control. While this methodology can be effective to ensure that a control strategy is operating as intended, it is a reactive, lagging approach to process understanding. Advances such as deployment of real time multivariate statistical process monitoring have helped to drive proactive approaches to detect weak multivariate signals within complex datasets and have been successfully utilized at Amgen to enhance monitoring controls to ensure robust performance, early detection of issues and to rapid root cause determination when process deviations occur.

The ideal future state however is to truly implement an automated data infrastructure to enable the capability to allow developers, data scientists, and users within manufacturing and PD to access a variety of structured and unstructured data sources from all stages of the process lifecycle to enhance monitoring, analytics and development of predictive process models. This presentation will describe Amgen's efforts to drive towards that ideal state through implementation of a data infrastructure that enables network wide data aggregation, predictive modeling, advanced process monitoring, and data science driven approaches to extract the knowledge within our big data.