#### 22 March 2021

**Problem solving session: analysis of batch process data from multiple sources for process improvement and for reduction in testing costs**

Dr. Jacqueline Asscher (Kinneret College and Technion), Yossi Mendel (Teva Pharmaceutical Industries)

This project was initiated by engineers and scientists who have been exploring the exciting combination of new online data and new methods of data analysis. We will share our characterization of the data they collected and the questions they are asking. After giving a very brief description of the methods tried so far and issues that have arisen, we anticipate hearing your suggestions for tackling these problems.

The data:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type of data | Example | Frequency | Availability | Cost |
| Batch level, inputs | Name of equipment; tests of raw materials; process settings;  length of process | Per batch (batches are ordered), e.g. 80 batches | Before or at start of batch, typically immediately, some at end e.g. length of process | Typically minimal, apart from additional testing e.g. of raw materials; might require additional data collection |
| Batch level, outputs. Properties of either the product or the process | Yield; impurities; length of process | Per batch (batches are ordered), e.g. 80 batches | End of batch, varies: immediately or after hours or days | Zero marginal cost |
| Offline within batch, outputs. Properties of the product | Impurities | Very low e.g. every 24 hours over 120 hours | After two hours | Moderate |
| Online within batch e.g. sensors | Temperature, pressure, pH, ammonia level | High, e.g. every 10 minutes over 120 hours | Immediately | Low |
| Offline within batch e.g. lab tests.  Properties of the product | pH, ammonia level | Low, e.g. every eight hours over 120 hours | After several hours | High |

The questions:

Question 1 – process improvement (e.g. increase yield, reduce impurities, shorten process length): How do we characterize batch level features of the within batch data (both online and offline) in order to generate batch level variables for use in modelling the batch level outputs?

Question 2 – reduction in testing costs: Could we replace some of the traditional expensive and slow offline tests (e.g. lab tests) with newer, cheaper and faster online tests (e.g. sensor readings)?

Question 3 – How can we obtain a short term, mid-batch prediction of the within batch process parameters, using both online and offline data?

**How have you used statistical methods to answer these three questions?**

**What other problems have you solved using this type of data?**

Methods tried so far and issues that have arisen:

JMP Pro’s Functional Data Explorer was used to address Question 1. Should we treat this analysis as an underlying characterization of the process, to be updated as future batches are added to a data base, or as specific to this sample of 80 batches? If this project leads to process improvement, will batches become more similar and the FDE tool less useful? How should we treat a change over time in the batch level variables?

JMP Pro’s Neural Nets were used to address Question 3. How can data from low frequency offline tests be included in models based on high frequency online tests?