# Context

Data-driven monitoring technqiues are useful for complex processes where corresponding first principle models may be difficult to develop while large amounts of meaningful process information is available in the historical data.

Multivariate statsticical process monitoring (MSPM) method is the most common.

Statistical monitoring for continuous processes (MVSP), using PCA/PLS and monitoring stats T2, SPE, Q

Batch processes are unique challenge because:

1. Transient behaviour
2. Highly non-linear and non-Guassian process behaviour
3. Batch-batch variations
4. Multi-phase behaviour

The argument for batch processes is that they are more necessary for cost-effective manufacturing of low-volume, high-value products that are unsuitable for continuous processes.

Typically followed two main strategies:

1. offline control – adjust the parameters of the next batch based on results of the current batch
2. online control – control charts for major process variables (problem – with many variables, they may all be within their control zones but compounded together leading to incorrect batches). OR compliance based – maintain the process as close to a ‘golden reference trajcetory’ as possible with the goal that it would logically result in on-quality product (problem – does not consider that batch-batch variations)

(1). MSPC, as applied to a continuous process in steady-state, would consider the signals as stastically stationary – the signals do not change with time. This requirement is usually not satisfied in batch/semi-batch processes. was overcome by Extended to batch processes with the pioneering work of Multiway-PCA (and later MPLS) by McGregor (1995) based on the fact of …

(2) Multiple improvements were developed to account for various factors such as dynamic characterics, nonlinearity, non-gaussian distribution etc. (k-PCA, h-PCA, d-PCA).

(3) batch-to-batch variations dictate

(4) traditional MPCA/MPLS models cannot efficiently capture the multiphase data behaviour that is common is most batch processes. These phases arise from multiple steps in a single processing unit, chemical or phenomenological (chemical reactions, microbial activities) which affect the underlying process correlations.

single monolithic models were found to be lacking. Each phase usually has its own characteristic dynamics and requires individual treatment. Variables have different significance and correlations in each phases which cannot be accurately captured by a single model. Furthermore, not all variables are even present over the entire course of the batch meaning inputs would have to be imputated with an obvious effect on performance.

Better solution was to use a multi-model approach. This means that the non-linear process can be split into multiple phases which conform to a mostly-Guassian distribution

and each model can have different inputs based on what data is available at that point in the batch evolution

### Evolution of Multi-model approach

The multi-model approach spawned various branches of consideration:

1. Phase detection – the main challenge in monitoring multiphase process is the number and manner in which it is divided
   1. Early work used specific expert process knowledge to logically divide the process
   2. Later, changes in key variables (‘indicator variables’ e.g. conversion, pressure) marked by Singular Points (e.g. discontinuities, inflection points, etc.)
   3. Modern approach use phase detection (in what is essentially a clustering exercise) to divide process into phases that are approximately linear. These include K-means, GMM. Generally these require the number of clusters to be known which again requires expert knowledge
   4. This is partially overcome with extensions such as finite-GMM, FJ-GMM, GMM-PSD, VB-GMM which automatically determine the number of phases during the EM phase.
2. Phase assignment – with the use of GMMs for phase division, calculating the posterior probability of a new sample data w.r.t. each phase and assigning the max is the golden standard for determining current phase.
3. Unequal batch lengths
   1. The end of a batch process is typically defined by reaching some goal as opposed to a set period of time. Due to variations in initial conditions, process conditions and operator intervention the length of batches is rarely ever the same
   2. Batch lengths have mostly been assumed the same in literature which is unrealistic
   3. Some techniques such as replacing time by some kind of indicator variable (e.g. %binder added, or 1% temperature increase) can be used if this is what is defined by the recipe…again requires process knowledge
   4. More recently DTW, specifically beta-DTW has been shown to be very effective at synchronizing batch
      1. But synchronize to what?
      2. A simple and effective approach – synchronize to a reference trajectory and assume the time domain of that batch. The reference trajectory can be chosen by a so called ‘golden batch’ which represent the ideal trajectory in optimial conditions. Or can be generated as the average of the historical data when this batch is unknown.
4. Transition-phase behaviour
   1. Owing to unequal length batches and variations in exact timing of key process events (e.g. switching from batch-> fed-batch) the level of uncertainty in phase-assignment increases in the transition region between phases leading to inaccurate predictions and/or false positives
   2. “deviations around key landmarks in the process stem from the fact that corresponding events occur at different times during batches. These need to be synchronized
   3. Some modern approaches include:
      1. Identifying transition regions explicitly as the overlap between adjacent models….local models can be built specifically for these transition zones
      2. Bayesian model averaging (BMA). Use the posterior probability of adjacent models as an adaptive weighting factor in the final prediction

# Project Goal

# Viewed as an extension of Wang, which implemented a quality control framework strategy similar to MCC incorporating multi-model, quality prediction and control actions similar to MPC.

<describe Wang approach in more detail>

<traditional compliance vs. closed-loop quality control approach>

<Shortcommings>

<what I plan to to do>

Take the control strategy of Wang, and extend it to account for:

1. No expert knowledge on the process required
2. Unequal batch lengths
3. Inter-phase transitions

# Approach

<description in words>

<diagram in 3 parts>

<the kind of process faults able to detect and fix with this..Wang/, 2018 paper (Liu), Proj. Proposal)

# Brief Literature reviews

<most relevant articles…3-4>

# Problems

Cannot reconcile the BMA inter-phase strategy with the control action optimization problem

Wang approach assumes initial conditions are known for each batch, not sure if I want to follow this but performance may not be good enough without it.

Wang requires that the initial conditions be know. (Liu,2018) reckons that the BMA is sufficient to “effectively handle the stochastic feature caused by process dynamics and batch-to-batch uncertainty but also ensure the accuracy of the assessment result”

I could probably just make the control action calculation with each model, then average the result by weighting of the posterior probability…only downside is increased computational load

Online phase-assignment – the GMM technique that im using is typically for differentiating between ‘modes’ of a process. That can be continuous processes shifting from ‘optimal’ to ‘sub-optimal’ or batch process from ‘product A’ to ‘product B’ production automatically. But the ‘modes’ that it is detecting in my process are just phases of the same recipe in which case I would always expect that the sequence of phase transitions remains the same.

DTW synchronizing…to what? Each other? What makes the most sense?

From research – two kinds of DTW…one creates a new time-scale, one assumes the time-scale of the other. It makes more sense to do the latter

1. Symmetric –DTW (minimize difference of two signals onto a new time-axis)
2. Assmyetric- DTW (maps the test signal onto that of the reference signal)

# Annexure A: HFO Specification