

# Batch Length Variations (Yiu)

Most of the initially added substrate is consumed by the microorganisms after about 45 h

switched from preculture stage to fed-batch stage which usually continues about 355 h

operating performance depends on the final penicillin concentration (final economic index)

50 training batches are produced under normal operating conditions (durations 390hr – 420h) by varying the normal operating ranges

Based on this, GMM-PSD defines 5 phases:

1. Pre-culture stage (45 hr)
   1. Phase 1 – [1-30]
   2. Transition – [30-34]
   3. Phase 2 – [34-43]
2. Fed-batch stage (355 hr)
   1. Transition – [43-47]
   2. Phase 3 – etc.

Even though pre-culture is only 45 hours, it detects 2 phases because of the rapid changes in variables

The fed-batch stage is 8x the length but only requires 3 phases because the variations slowdown

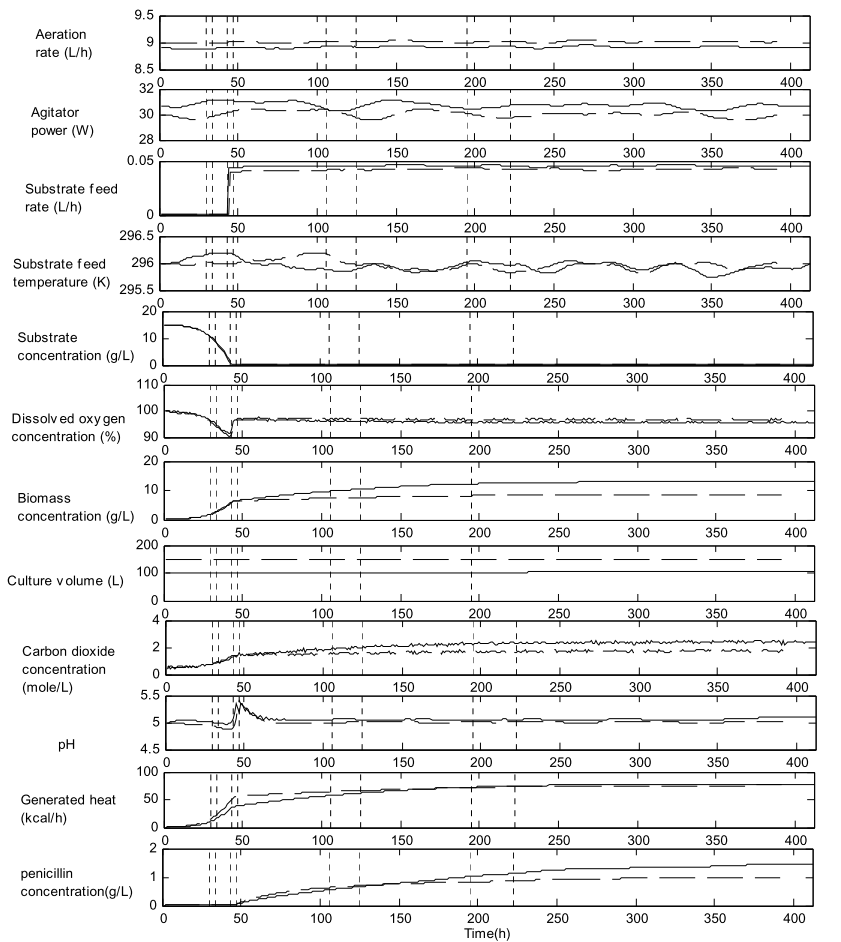
The transition intervals are from batch-to-batch variations (purity, etc.) which cause unequal length batches.

## Optimal vs. Non-optimal

two test batches under optimal and nonoptimal operating performances are generated separately.

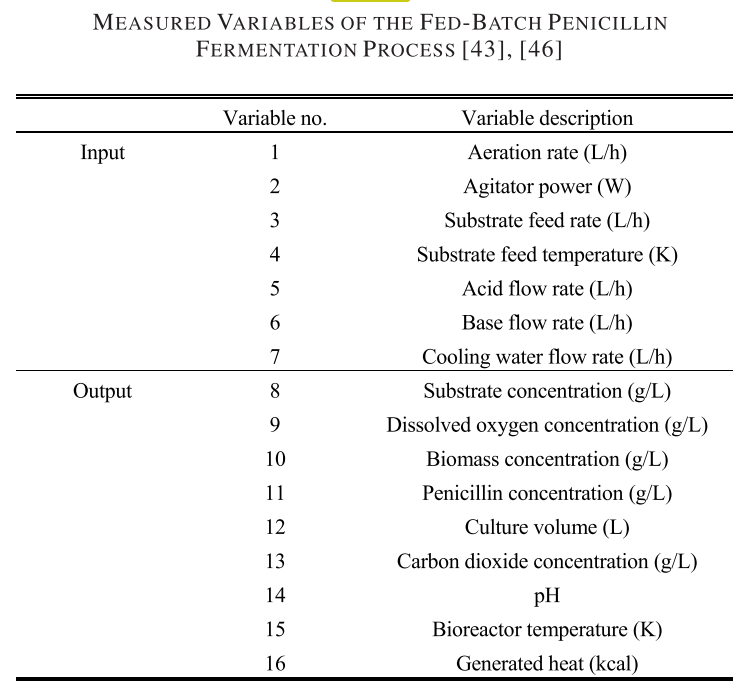
In optimal test batch, the aeration rate, agitator power, substrate feed rate, pH, and culture volume are in their optimal ranges as 8 shown in Table 7. Its duration is 412 h.

It is known that the volume loss due to evaporation is very significant for penicillin production since you lose biomass and substrate. The non-optimal test case is done using a culture volume of 150L (within normal operating range but on the low end). The duration of the nonoptimal test batch is 392h.



# Fixed Length Batches (Jiang)

Sixteen variables are considered and listed in Table I.



The entire duration of each batch is 400 h. (1hr sampling interval)

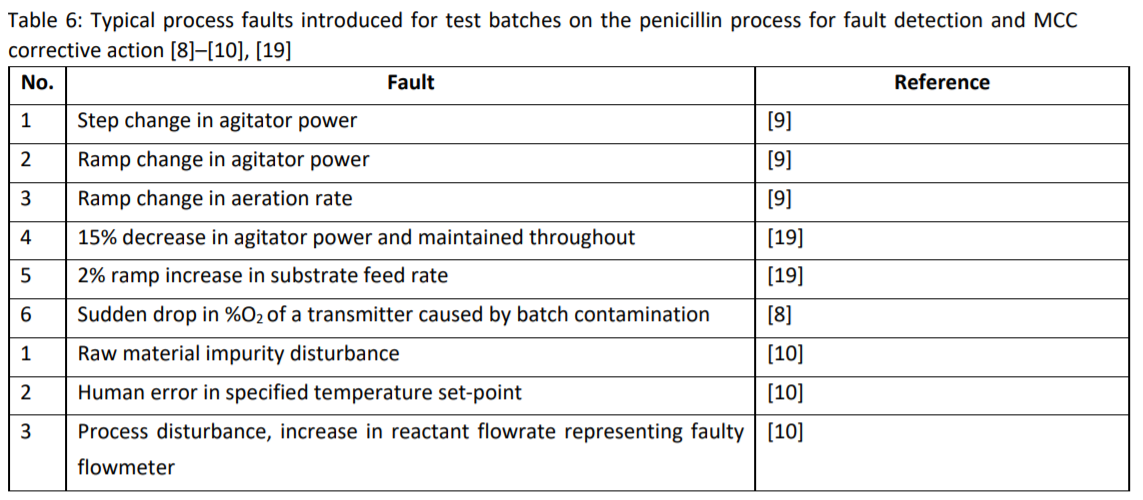
The initial conditions and set points of the simulation are presented in Table II.

A total of 100 batches are simulated under normal operating conditions, and the process data are collected as training data.

**Fault 1,** a step change is introduced to agitator power from the 150th to the 300th point.

**Fault 2**, a ramp change is introduced to agitator power from the 150th to the 300th point.

**Fault 3,** a ramp change is introduced to aeration rate from the 150th to the 300th point.



* **Unequal batch lengths -** Since batch process termination is defined by quality as opposed to time, the overall length of batch data and the timing of key steps/phases must be synchronized

The multiphase behaviour and variations in batch length often give rise to complex nonlinear behaviour in the transition regions where assignment to a specific local phase becomes difficult [1].

### 3.3.2 Synchronization

termination of a batch is typically fulfilled by having achieved an acceptable final product quality and, considering the nature of batch-to-batch variations, non-linearities and other features inherent to batch processes, this often results in historical data where batch lengths are unequal [5].

Since considerable statistical information regarding the correlations between input and output variables are captured in the trajectories it is required to align and synchronize the data before implementing a process monitoring framework.

This step becomes significantly more crucial in batch process with multi-phase behaviour since the unaligned intervals will lead to mischaracterization [19].

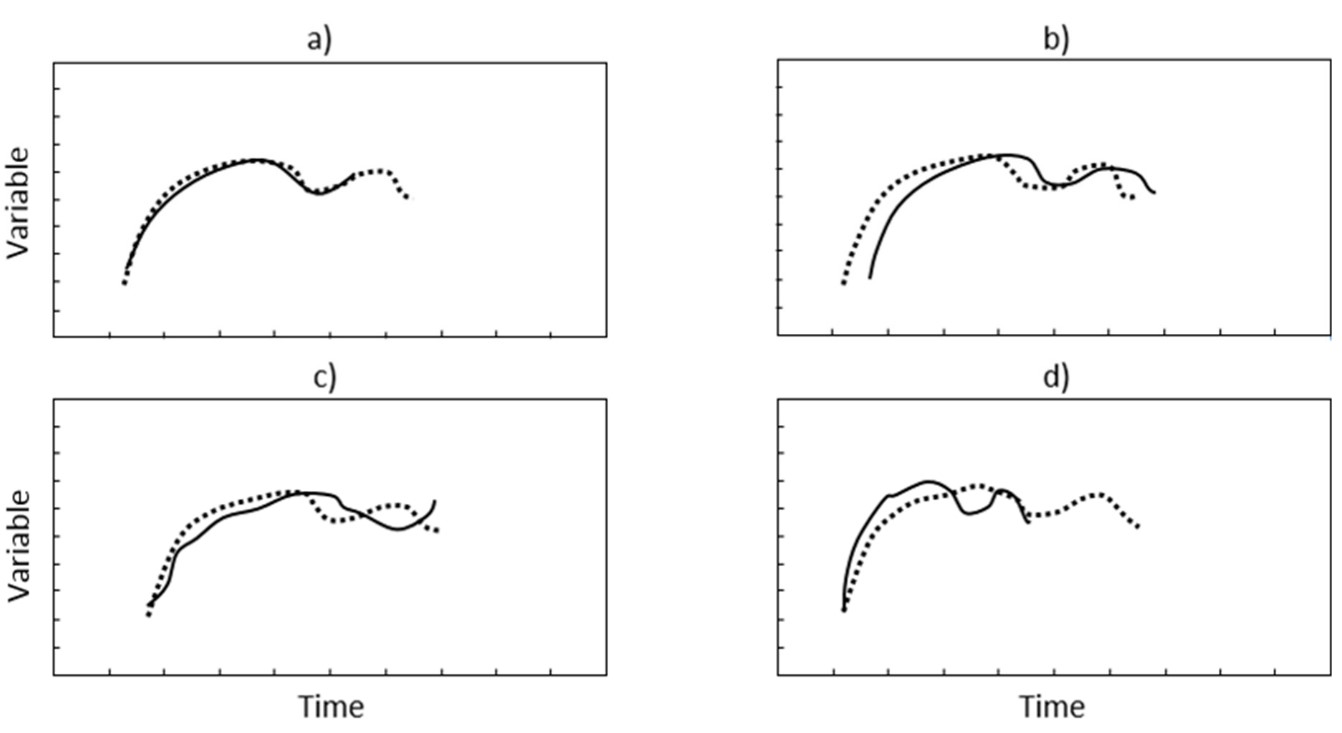


Figure 2 – illustration of two unequal, unsynchronized trajectories of the same process. Redrawn from [5]

Several techniques exist the most popular of which is **Indicator Variable technique** (IVT), **Dynamic Time-warping** (DTW).

In IVT, a variable is chosen to represent the maturity of batch and major milestones. The milestones correspond to a percentage of the IV at the current time point with respect to its final value and observations are aligned relative to this.

Dynamic Time Warping can synchronize trajectories with translation, stretch or compression applied offline found in various studies [4],   
It aims at aligning two sequences of feature vectors by warping the time axis iteratively until an optimal match between the two sequences is found.

To Reference trajectory or all to each-other?  
  
with several variations such as Relaxed Greedy Time Warping (RGTW) [42] and Extrapolative Time Warping (XTW)[4].

## 3.4 Phase Division

In general, three different approaches for phase detection are employed, (i) expert knowledge, (ii) process analysis, and (iii) data-driven [19].

**Expert knowledge**

typically corresponds to phases based on operational stages or intrinsic changes in underlying activity (e.g. microbial growth stages) using prior knowledge.

A common technique is the use **of singular points (SP)** in key variables which are points of departure (discontinuities, inflections, minima or extrema etc.)   
Phase Changes are then synchronized to a reference ‘golden’ trajectory for monitoring [4].  
[11] Characterized fermentation phases based on SPs in the pH and substrate federate  
[12] similarly used dissolved oxygen in fermentation process for phase identifications.

**data-based methods**

phase is defined based on underlying statistical correlation of data where significant departures are indicators of consecutive phases.

Do not rely on prior process knowledge

Essentially a clustering exercise

[19] used automatic ISODATA dynamic clustering (similar to k-means algorithm) to detect phases followed by DTW for synchronization of these unequal sub-phases

* + ISODATA does not require you to specify number of clusters like k-means does
  + The duration of each batch is 400h, comprising a pre-culture stage (about 45h) and a fed batch stage (about 355h).
  + 40 **batches** are generated - Small variations were added to the simulation input data to mimic the variations in the normal operating conditions
  + 4 phases automatically detected
  + Each sub-phase undergoes recursive DTW to synchronize before model development

STM-PCA = automatic phase division with transition zones based on changes in loading matrix

MP Algorithm = automatic phase division based on regions that agree to a linear model

**Clustering Methods:**

[22] K-Means / GMM = require number of phases to be specified a priori

Variations

Finite GMM, Variational Bayesian GMM, GMM-PSD etc. = use some algorithm to automatically specify number of phases