Identifying highly variable and energy intensive batch manufacturing processes using statistical methodologies

Konrad MULRENNAN[a,1], Mohamed AWAD[a], John DONOVAN[a], Russell MACPHERSON[b] and David TORMEY[a]
[a] *Centre for Precision Engineering, Materials & Manufacturing Research, Institute of Technology Sligo, Ireland*
[b] *GlaxoSmithKline, Sligo, Ireland*

Abstract. Batch manufacturing processes are extremely energy intensive. These processes can benefit from statistical analytical methods to identify production phases that are highly variable or contain outliers. Outliers, in this sense, represent batch processes that take significantly longer time to complete, which in turn results in the consumption of much greater energy than other observed identical batch processes. This work presents a case study on the analysis of manually recorded written batch manufacturing records from a pharmaceutical facility. As batch records are currently recorded in writing, it poses a barrier to rapid identification of process variability and outliers. The benefit in identifying processing steps that are highly variable is the introduction of standard operating procedures that reduce variability. Outlier identification allows for batch processes to be further investigated so that root causes are identified and acted upon to prevent future occurrences. The highly variable process steps and outliers are identified using boxplots. These are further analysed to identify causes, which include transcription error in data recording, parallel processing between manufacturing locations sharing utilities and heuristic based decisions made by plant process technicians. By identifying and acting upon these causes, the facility can achieve greater energy efficiencies and have a more sustainable approach to batch manufacturing.

Keywords. Batch manufacturing, data analysis, sustainable manufacturing

1. Introduction

There has been significant drive towards more sustainable manufacturing processes for a number of years [1]–[3]. A key metric in achieving more sustainable processes is reducing the energy used in manufacturing [1], [3]. Energy reduction cannot be achieved until there is a sufficient level of instrumentation and monitoring [4]. Shrouf *et al.* discuss that awareness of energy consumption needs more granular monitoring and also be linked to production data to achieve reduction and minimisation of energy consumption in manufacturing [5]. The mapping of energy data to production processes can prove challenging when a factory is not instrumented to the level of a smart factory and manufacturing records are recorded manually and handwritten.

[1] Corresponding Author. mulrennan.konrad@itsligo.ie

The following work presents a simple solution for the rapid identification of highly variable manufacturing cases that in turn exact significantly more energy demand than normal cases. This is achieved through the use of statistical boxplots [6]. These are used to analyse the batch manufacturing records of a case study pharmaceutical company. By identifying the highly variable manufacturing cases and taking action to reduce the variability in product manufacturing time or phase completion time, will in turn, result in a reduction in manufacturing energy demands. An aim of this research is to minimise the plant energy consumption of the facility. To achieve this, the energy demands of the facility must be fully understood. This work proposes a method to utilise simple statistical techniques to gain a more complete understanding of the manufacture of all products in the facility.

2. Case study

GlaxoSmithKline (GSK), a world leading pharmaceutical health care company, currently operate a dermatological batch manufacturing facility in Sligo, Ireland. Skincare products such as creams, lotions, gels and liquids are produced in large volume manufacturing vessels in this facility. Highly variable production phases for each product can exact significant demand on plant energy capacity. Identification of product manufacture and production phases that experience significant variance in the time taken to complete, is key to introducing measures to achieve more sustainable practices. Batch manufacturing data records have been digitised from hand written documents for approximately 1000 batch manufacturing cases. These records represent all batch manufacturing cases over a one year period from April 2017 to April 2018. As the original recordings have been hand written, there are a number of possible transcription errors that may occur and are only evident when analysing the time series. These include the time being recorded in a 12 hour or 24 hour format and date entry produced errors by shift workers who begin a shift before midnight and end their shift the following morning. These transcripts are easily interpreted by a human reader.

The challenge in analysing the digitised records is that the time series recordings become transcription errors if they do not share an identical time format or the datetimes do not follow a logical order. Datetimes are the combined date and time assigned to an event. The datetimes may not follow logical order if there are date entry errors as previously mentioned or through the different heuristic decisions of the technicians who control the process and are recording the information. In the second case, an engineer may begin a phase prior to the completion of a current phase e.g. a mixing phase may begin while a heating phase is ongoing until the set temperature is reached. This can result in the sequence of events not following a logical order as heuristic decisions can differ between process operators. In addition, each product that is manufactured on location has a different recipe and different manufacturing process steps. This requires that bespoke analytical programming scripts be written for each product, as a single automated script for all products cannot be achieved. As soon as the analysis is completed for each product, it results in a rapid identification of manufacturing cases or manufacturing phases that are highly variable and result in significant energy demand for the facility. Once these have been identified, measures can be put in place to prevent future occurrences.

2.1. Data preparation and analysis

The hand written manufacturing records were digitised into a tabular spreadsheet format for each product. Each recorded event was assigned a datetime, which was a logical decision that coupled the manufacturing start date with the event time or the manufacturing end date with the event time. At an individual level, each product's manufacturing phase was identified and the time taken to complete each phase was calculated. Once this stage was complete for a given product, boxplots that presented the total manufacturing time, the time taken to complete process phases and the times between process phases were generated. A manual troubleshooting process assessed each boxplot, in consultation with GSK production personnel, to identify the transcription error cases that have been described. The outliers of each boxplot were investigated in the data set for time format errors, date entry errors and phase logical ordering errors. The transcription errors were resolved once identified by correcting the time format or date. The boxplots also allow for rapid identification of highly variable process phases and also the identification of batches that are significantly longer than all of the others at a particular production phase. These are the remaining outliers in each boxplot that are not transcription errors.

3. Results and Discussion

Figure 1 presents the results of the outlined methods for a sample product that has been labelled "Z". Each manufactured batch has been labelled in a random order with letters from the alphabet. The y-axis presents time and this has been normalised. Each of the process phases are identified on the x-axis. P1 represents the total time of manufacture of each batch of product "Z". Each of the subsequent phases P2 to P8 represent the logical order of the manufacturing process from beginning to end. The phases P2 to P8 have been anonymised and represent heating, cooling and mixing cycles and the times between those cycles.

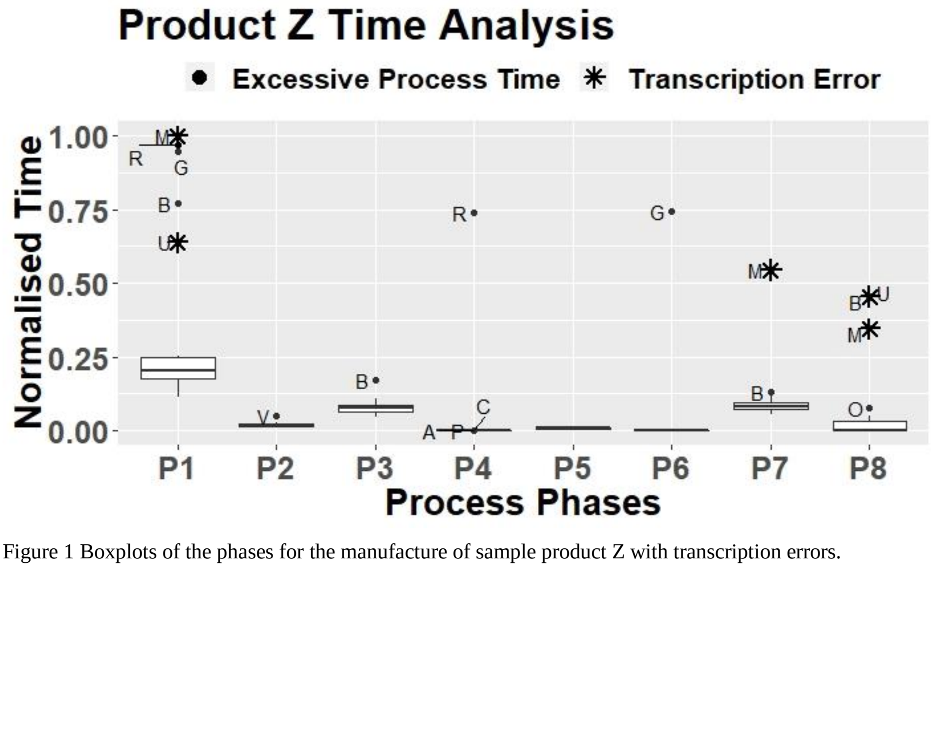

Figure 1 Boxplots of the phases for the manufacture of sample product Z with transcription errors.

The “✳” points (M, U) in Figure 1 represent transcription errors as have been identified by the manual troubleshooting process. Once these have been identified the errors are corrected and the updated analysis is presented in Figure 2. The outliers in Figure 2, which are represented by black dots, are batch manufacturing cases or manufacturing phases that take significantly longer to complete than a general batch manufacturing case. In Figure 2, there are three batches (R, G, B) that have been identified as outliers for P1 (P1 represents the total time of manufacture, which is cumulative of all other phases P2 to P8). By looking at each of the individual manufacturing phases (P2 – P8) it is very easy to identify where the excessive times in manufacture occurs for each batch i.e. R has excessive time in P4, G has excessive time in P6 and B has excessive time in P3, P7 and P8. This makes it very easy to focus effort on identifying what was happening during those process phases so that the root cause can be understood and eliminated for future batches. Other outliers are identified occurring only at specific phases.

Once outliers are identified, the information relating to the batch process or phase is located in the digitised manufacturing record. This allows the cause of the delay in completion of the overall process or the individual phase to be identified. When this information is known it can be acted upon. The current investigative procedure is very manual based, time consuming and hugely challenging in respect to gathering the data and performing the analysis for gaining any meaningful insights. That poses great challenges to react to cases where there is a significant difference in the manufacturing time for the overall process or phase. The cause is simply that these cases cannot be identified easily. The proposed method resolves this issue if data is collected digitally.

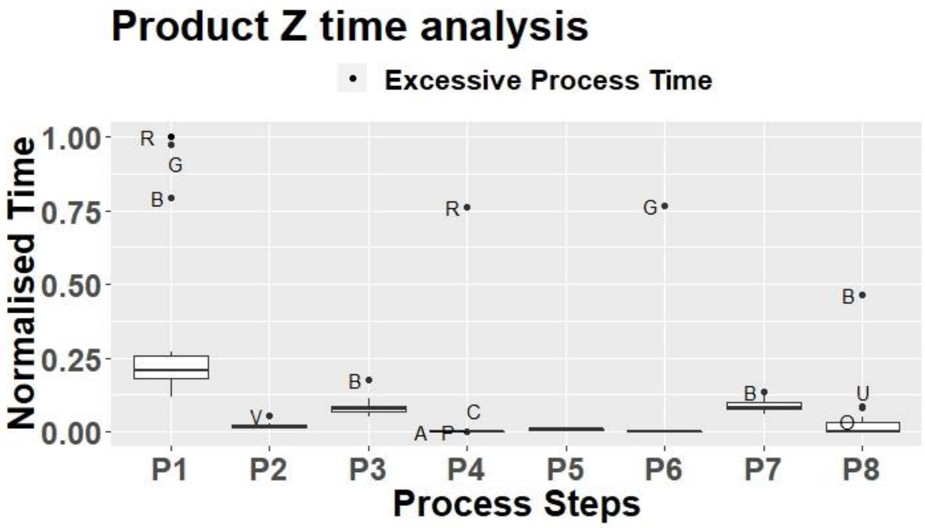

Figure 2 Boxplots of the phases for the manufacture of sample product Z without transcription errors.

Boxplots are useful tools as they provide measurement information (time in Figures 1, 2 and 3) and also provide information related to the distribution of the measurements. The distribution of the measurements also identifies the phases that are most variable, as presented in Figure 3. Figure 3 has zoomed in on the y-axis to present the distribution of time measurements for the manufacturing process and phases succinctly.

The outliers in phase P4 (A, P, C) can be ignored as these are non-zero values recorded for that phase that are very close to zero. Phases P3, P7 and P8 have a large distance between the minimum and maximum values. This indicates that general process improvements may be possible as the minimum values for each phase are indicative of the minimum required time to complete the phase. Phases P2, P7 and P8 present skewed data distributions, which indicates that a number of the batches are completed in a consistent amount of time while the rest of the batches are completed in more variable amounts of time. Possible reasons for this variation are that the manufacturing locations share utilities and also the heuristic decisions of process technicians. In the first case, it is acceptable to consider that when a number of products are being manufactured in parallel there will be considerable demand on the shared heating and cooling systems employed. This increase in demand will inevitably result in longer and more variable times to reach temperature set points. The heuristic decisions of the technicians will take into consideration all manufacturing cases with a target of optimisation. It is difficult to gain insight into these cases as the decisions and reasons for the decisions are not recorded.

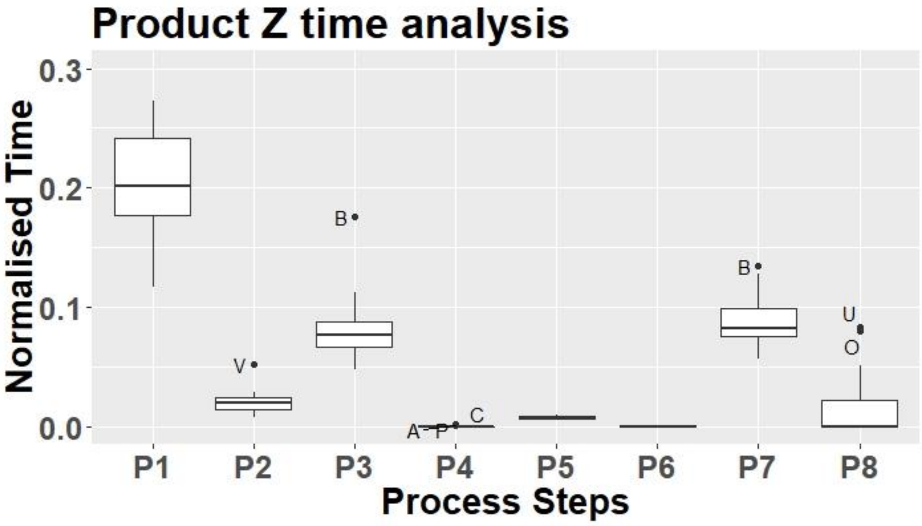

Figure 3 Boxplots of the phases for the manufacture of sample product Z to investigate variability of phases. The y-axis has been zoomed from Figure 2 to view the boxplot distributions succinctly.

Introducing an electronic recording system would eliminate all of the problems discussed and present real time aggregated analysis that could instantly identify outliers and alert process technicians. Additionally, systematic changes to operating procedures can be implemented when highly variable production phases have been identified. The cause of the high variability can be determined and once known these can be acted upon. Reducing variability will inevitably lead to more sustainable manufacturing practices and energy reduction. This will be achieved as the manufacturing time for each batch and the time needed for each phase will become more predictable. This will make it easier to manage and schedule production as the demand for shared utilities can be predicted in advance of manufacture. Introducing better operating procedures that incorporate the best practice heuristic decisions of the process technicians will also contribute to a reduction in process variability. It is quite common for best practice to be known only to the experts who are inevitably involved in the batch manufacture of the products every day. Other studies have revealed a difficulty in transferring heuristic knowledge into computer software [7]. The approach described here allows for best practice to be identified through analysis of best, worst and general batch manufacturing cases.

4. Conclusions

The total time of production and the time for each phase of production for all products can be variable between manufacturing runs. This variability is caused by a number of factors. The manufacturing batch records are hand written. This method of recording data introduces a risk of transcription errors and this is further compounded when these records are digitised for further analysis. All production utilities are shared, which results in significant delays for production phases to be completed if multiple processes or manufacturing locations are operating in parallel. Typically, manufacturing process technicians make heuristic decisions related to production scheduling and determine when production phases are complete. These heuristics are based on significant experience and implicit knowledge built over years in the sector and are difficult to transfer to a formalised knowledge base. The manual troubleshooting process is very time consuming and provides an obstacle to the identification of excessive time to completion of energy intensive processes. The simple approach that has been discussed in this research work can deliver real time analysis if an electronic data recording system is employed. This could potentially save on costs and energy resulting in more sustainable manufacturing practices for batch manufacture in the pharmaceutical sector.

Acknowledgements

The North West Centre for Advanced Manufacturing (NW CAM) project is supported by the European Union's INTERREG VA Programme, managed by the Special EU Programmes Body (SEUPB). The views and opinions in this document do not necessarily reflect those of the European Commission or the Special EU Programmes Body (SEUPB). If you would like further information about NW CAM please contact the lead partner, Catalyst Inc, for details.

References

[1] F. Apostolos *et al.*, "Energy efficiency of manufacturing processes: A critical review," *Procedia CIRP*, **7**, (2013), 628–633.

[2] M. Despeisse *et al.*, "The emergence of sustainable manufacturing practices," *Production Planning & Control*, **23**, (2012), 354–376.

[3] H. A. ElMaraghy *et al.*, "Energy use analysis and local benchmarking of manufacturing lines," *Journal of Cleaner Production*, **63**, (2017), 36–48.

[4] C. Jiménez-González *et al.*, "Evaluating the 'Greenness' of chemical processes and products in the pharmaceutical industry-a green metrics primer," *Chemical Society Reviews*, **41**, (2012), 1485–1498.

[5] F. Shrouf *et al.*, "Multi-level awareness of energy used in production processes," *Journal of Cleaner Production*, **142**, (2017), 2570–2585.

[6] J. W. Tukey, *Eploratory Data Analysis*. Addison-Wesley, 1977.

[7] H. Stefansson *et al.*, "Discrete and continuous time representations and mathematical models for large production scheduling problems: A case study from the pharmaceutical industry," *European Journal of Operational Research*, **215**, (2011), 383–392.