API production process improvement project: a Six Sigma approach

Bernardo Maria Fernandes Ferreira
bernardo.maria.ferreira@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

November 2021

Abstract

This thesis aimed at the improvement of a generic API production process at Hovione´s Sete Casas site. Yield optimization of the final product production step was the problem to solve through the DMAIC cycle methodology. Over the analysed period (from July 2018 to January 2021), yield had an average value of 81.83% with a relative range of 11%. It was accounted that, on average and per year, a full batch throughput is lost due to yield variability. Multivariate data analysis techniques were used in order to find statistical correlation between input quality attributes and process variables with the response variables. Impurity H and G present in the input material to the final product process step were found to negatively impact the yield. The statistical analysis of the process variables revealed that the crystallization is the most critical to yield operation. The models were re-built considering the final product´s assay instead of yield as response variable. The contribution of the process variables to the quality of the product was on the same line as for yield. The process leading to the input intermediary of the final product process was also analysed taking impurity H and G as response variables. Although data was only available for 6 production batches, some actions could be taken from the models. The improvement actions were screened based on their impact and effort and an interactive control sheet as well as a summary flowchart of the generated process understanding were elaborated in order to maintain the improvements.

Keywords: Pharmaceutical industry, API production process, DMAIC cycle, Multivariate data analysis.

1. Introduction

1.1. Motivation

The pharmaceutical industry has the noble mission of improving the quality of human life.

The *International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use* (ICH) quality guideline Q8 (ICH Q8) defines quality in the pharmaceutical industry as "the suitability of either a drug substance or drug product for its intended use. This term includes such attributes as the identity, strength, and purity" [1]. Further consideration can be added to this definition in terms of reliable clinical performance: a quality drug product "delivers clinical performance per label claims and does not introduce additional risks due to unexpected contaminants" [2]. Integration of these two approaches is done considering that the clinical parameters that are crucial to good clinical performance are derived from the quality attributes of the drug product or substance [3].

Since the *U.S. Food and Drug Administration* (FDA) report publication *Pharmaceutical Current Good Manufacturing Practices for the 21st Century* [4] on 2004, the industry´s approach to quality began to change from Quality by Testing (QbT) to Quality by Design (QbD) [5, 6]. In the traditional paradigm, quality is assured by a series of testing on raw materials and the final process output. Only when all specifications are met can the product be released to the market or proceed to the next step on the value chain [3, 5]. When all the specifications are not met, the batch has to be reprocessed or can even be discarded, leading to failure in meeting customer demand. It is estimated that, in the early two-thousands, 5 to 10% of the total batches produced in the industry needed reprocessing or were discarded [7].

Contrary to this traditional notion, comes the approach developed and coined by the quality pioneer, Dr. Joseph Juran, Quality by Design (QbD). This systematic, scientific and holistic perspective assures product quality on the design of the process thus eliminating the need for extensive testing on the final product [3, 8, 9]. ICH Q8 defines QbD as "a systematic approach to development that begins

with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management" [1].

The recent pursuit for quality in the pharmaceutical industry was combined with a quest for productivity increase that translates into effective use of the company's resources. Lean and Six Sigma have proven to succeed on the matter and although slower than other industries [10, 11, 12], began to gain ground on the turn of the century in the top pharmaceutical companies [11, 12].

The successful integration of quality and productivity, based on a scientific understanding of manufacturing processes, is, nowadays, on the top of the agenda of pharmaceutical companies [13, 14] paving the way for operational excellence (OPEX) programs in an attempt to manage cost, quality and time while at the same time focusing on customers needs [15, 16].

The need for a continuous improvement culture in the pharmaceutical industry, that relies on process understanding [17], is evident based on the number of methodologies and programs that have been launched and pushed by regulatory agencies and applied (or are still to be) since the beginning of the 21st century [18].

1.2. Topic Overview

Corticosteroids are a class of molecules produced naturally on the adrenal gland (located above the kidneys) that have a direct impact on stress and immune response, protein and carbohydrate metabolism, blood electrolyte levels, the regulation of inflammation, and behavior [19]. Since their discovery, corticosteroids have been used in almost every area of medicine and administered by nearly every route [20]. They are one of the most prescribed classes of drugs worldwide with an estimated 10 billion dollars per year in sales [21].

Fluticasone propionate is a corticosteroid that can be administrated via oral, nasal or topical route [22]. The route of administration depends on the condition to treat [22, 23, 24]. Hovione holds the patent, producing it on Sete Casas installations by means of a process developed in its R&D center. The process can be divided into 5 steps, giving each an isolated intermediate as output.

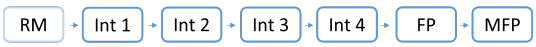

Figure 1: Process overview displaying the several steps that lead to the final API product.

The API (further referred to as FP) is produced by campaign on a multipurpose installation (also intended for APIs of the corticosteroids family). General metrics concerning the production of FP are given in table 1.

Table 1: Overview metrics for all the steps of the production process of the FP on a time frame from 2018 to 2020. The standard deviation is also presented.

Process step	**Avg. yield (%w/w)**	**Expected yield (%w/w)**
Int 1	94 ± 0.6	92 ± 5
Int 2	102 ± 2.6	99 ± 8
Int 3	106 ± 2.2	107 ± 5
Int 4	103 ± 2.4	101 ± 10
FP	78 ± 15.6	70 ± 28
MFP	94 ± 2.9	95 ± 5

The yield of the production steps can be viewed as the throughput ratio. To better understand this variable (%w/w), the following formula should be considered:

$$Yield(\%) = \frac{Net\ weight\ obtained}{Net\ weight\ loaded} \times 100 \qquad (1)$$

The formula used for yield calculation does not consider the purity of the final substance obtained nor the changes in molecular weight of the isolated intermediates or final product. As such, the ratio obtained can be higher than 100%, which should not be considered abnormal.

The uncertainty in the expected yield is increased moving up on the production train and stopping on the final API before size reduction (which happens from FP to MFP). This increase is roughly accompanied by an increase in the standard deviation of the average obtained yield since 2018. High variability in the yield obtained ultimately leads to ineffective use of the production resources (equipment, personnel, utilities, etc) and so, in accordance with the production team, this was the primary problem to target.

The work methodology followed a backwards approach. The problem is identified on the process output and the cause(s) for the problem are searched firstly on the process that leads to the output and so forth, going backwards. In this way, the causes for variability on the output parameter will be spotted, improvement actions will be launched and tested. Process understanding will be further developed that leads, due to the implemented improvement actions, to an increase in process robustness.

2. Background

2.1. Six Sigma

The term "Six Sigma" originated from the very core of the philosophy itself: reduce variability to improve quality. Assuming that a random process

variable (X) follows a normal distribution with a designated mean, μ, and standard deviation, σ, then $X \sim \mathcal{N}(\mu, \sigma^2)$. The probability density function is given by the following equation:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})} \quad (2)$$

The percentage of data that lies within $\mu \pm 6\sigma$ is 99.9999998%. If the lower and upper specification limits for a certain process or product parameter (LSL and USL) are located at $\pm 6\sigma$ from the mean, then the proportion of defectives, *i.e.* the proportion falling outside the specification limits, would be 0.002 ppm. Allowing for a 1.5σ shift on the mean, one would get 3.4 ppm defectives (99,9996% within specification) [25]. This is the metric that the pioneers of Six Sigma at Motorola set out to achieve on all of their processes reflecting the goal of near perfection in terms of quality [25].

2.2. DMAIC cycle

Every process can be defined, measured, analysed, improved, and controlled (DMAIC). Six Sigma views all work as processes and so, all work can be defined, measured, analysed, improved, and controlled [26]. This sequence of actions is at the very basis of Six Sigma, the DMAIC improvement cycle or problem-solving strategy. Underlined in this problem-solving strategy is the simple equation [27, 28]:

$$Y(CTx) = f(X - influencers) \quad (3)$$

Measurable parameters have to be defined on the process output that can quantitatively describe the problem - these are the critical-to-x (CTx) variables of the project. Here the term "x" means any area that has an impact on the customer. Some examples of areas that are often subject to problem-solving projects are: quality (CTQ); cost (CTC); delivery (CTD); safety (CTS) [26]. These response parameters are determined by a set of variables, the X-influencers. If these influencing variables (X-influencers) are controlled, then the process outputs parameters are controlled as well.

2.2.1 Define Phase

A problem is measured on the output and can be defined as "an undesirable situation which may be solvable by some agent although probably with some difficulty" [29]. The first step of the DMAIC problem-solving cycle fundamentally aims to answer the two following questions:

1. What is the problem?
2. How big is the problem?

Together with the strategic goals, these two questions form the project statement that should help the project team focus on the core issues and establish a common starting point [26].

2.2.2 Measure Phase

The second phase of the DMAIC cycle deals mainly with in-depth process mapping and data collection. Batch processes originate data that can be arranged as a 3-way data table as illustrated in figure 2.

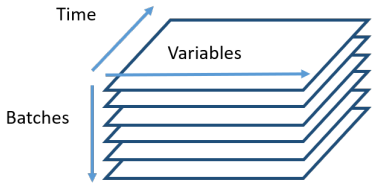

Figure 2: Three-way batch process data table.

For chemical synthesis processes in the pharmaceutical industry, *i.e.*, batch processes, two categories of data can be defined: process photograph and process film. The process photograph data is represented by the front face of the cube where to each completed batch a single value of a certain variable is attributed. It does not give a complete picture of the batch since it only displays a shot of the process and therefore cannot be considered for robust improvement actions derived from the analysis. The process film data gives a more complete overview of the process. It consists, for every batch, of the data time points of variables such as temperature, pressure, speed of the agitator, and pH among many more.

2.2.3 Analyse Phase

During the Analyse phase, the data collected during Measure is statistically analysed. It is during this phase that process understanding is consolidated: the X-influencers that mostly impact the identified Y parameters are signaled and through correlation analysis it is understood how does the variation in X influences the behavior of Y.

2.2.4 Improve Phase

The goal of the fourth phase of the DMAIC cycle can be divided into three consecutive parts.

The first one is to successfully materialize the results of the statistical analysis conducted on the previous phase into tangible, concrete, and feasible actions for problem resolution and process improvement.

Secondly, is idea prioritization. Usually, the effort to implement the actions that come out of the Analyse phase surpasses the time or resources available

to implement them and so, prioritization is imperative [30]. This action prioritization can be done by placing them on an Impact Vs. Effort matrix.

Still under the second objective of the Improve phase, an action plan with feasible timelines and accountable people for each task is drawn and is to be used in the third objective of the Improve phase: action plan implementation. Only on this stage is the process actually improved and the *status quo* is changed into a revamped version of the process.

2.2.5 Control Phase

Maintaining the improvements achieved is a part of this final phase of the DMAIC cycle. A rigorous process of documentation of the lessons learned during the entire project should be performed and a clear identification of how the improvements can be replicated and applied to other processes [26]. An often-used practical way for the improvements sustain is the development of training materials in order to ensure continued support for the people involved with the process on a daily basis.

3. Results

3.1. Define

3.1.1 What is the problem?

According to table 1, the expected uncertainty on the yield of each step leading to any intermediary in the API production train is increasing moving from the starting raw material to the final product. The increase in the uncertainty of the predicted yield is accompanied by an actual increase in the standard deviation of the obtained yields. High variability on any process output parameter can be translated into a poorly controlled process and lack of robustness, where process robustness can be defined as the lack of sensitivity of the process outputs to fluctuations in the process inputs and process variables [9].

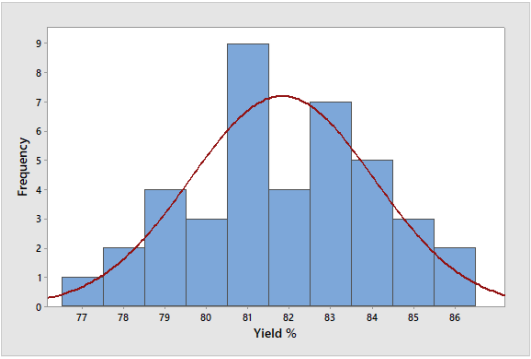

Figure 3: Histogram of the yield of FP process from July 2018 to January 2021 with normal distribution fitting.

A process in which the throughput is not predictable leads to an unknown number of batches needed to satisfy a client´s order that is inevitably accompanied by biased and uncertain production planning. High variability in yield also leads to ineffective use of the company´s resources since the cost of equipment, personal, raw materials, and utilities do not change according to the throughput. By standardizing and preferably optimizing the yield of the final step, the process advances to a state of tighter control, the throughput is increased and the company resources are used at a higher utilization rate. As side (but desirable) effects of the success of the improvement project, process understanding is gained and a culture of continuous improvement is fostered among the company.

3.1.2 How big is the problem?

In order to correctly calculate the impact of the project, a holistic metric has to be identified. As stated in the previous section, high variability on yield causes biased and uncertain planning that ultimately leads to missed opportunities in terms of throughput that can then be converted to missed opportunities in terms of revenue considering an average price of the final product and that on average, 17 batches of FP are produced per year.

The calculation for the possible impact of the succeeded project was performed through the following formula:

$$MO = \frac{\sum_{n}^{\infty}(SP - Yield_n)}{n} \tag{4}$$

where SP designates the optimization set-point considered (all batches on the time frame with the same yield) and MO the calculated missed opportunities.

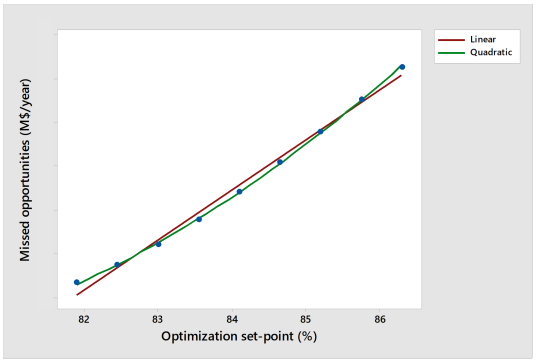

Figure 4: Graphical representation between the averaged missed opportunities in terms of revenue per year and the optimization set-point. A linear model ($R2$=99.15%) and a quadratic one ($R2$=99.97%) were fitted.

Although with a very small difference in the values of $R2$, the quadratic model displays a better fit to the data, showcasing that the missed opportunities in terms of revenue are exponentially dependent on the yield optimization. Towards the higher end of the optimization set-point, the possible gains will be higher than those of the lower end.

3.2. Measure

During the second phase of the process improvement cycle, a better understanding of the problem and of the process is done through process mapping. The IPO diagrams of each step of the production train were drawn. Regarding data collection, both process photograph and process film types of data were collected.

3.2.1 Process Description

The present project leaned over the analysis of FP and intermediary 4 production steps and so the response variables that will be present in results regarding the Analyse phase will be the yield of FP step and material attributes of intermediary 4. Only these two processes will now be subject to an overview description due to confidentiality reasons.

Intermediary 4

Intermediary 3 is dissolved and two inorganic salts are added. A final and gaseous reactant is added and reaction takes place. Multiple degasing steps take place in between the load of the several reactants. Precipitation happens due to antisolvent addition and cooling. The suspension is then filtered and dried.

FP

Intermediary 4 is dissolved. Precipitation happens through solvent evaporation, antisolvent addition and cooling. The suspension is then filtered and dried.

3.3. Analyse

This phase of the DMAIC cycle covers all the statistical analysis of the collected data. As stated before, multivariate data analysis techniques were used as a more holistic, robust, and feasible approach to univariate statistics. The results will be presented following the backwards approach: starting on the last production step and moving in reverse on the production train.

All the models that will be presented during the following section are termed black-box models or statistical models. Within this framework, systems are only viewed in terms of their inputs (stimulus) and outputs (responses), without any knowledge of their internal workings, and are built based only on historical or experimental data [31]. Contrary to white-box models, or mechanistic models, which are entirely based on mathematically expressed universal natural laws, black-box models only need the specification of the system´s inputs and outputs and are particularly useful when the system is poorly understood [31] which is the case for the chemical-pharmaceutical industry, where systems are often too complex and the simplifications made to achieve a mechanistic model often compromise and surpass the advantages for this type of modeling.

3.3.1 FP process step analysis

Quality

The yield on the final step (the response variable considered) was modeled against the quality data of the starting raw material of the final step (intermediary 4). Two impurities of this intermediate were removed from the analysis due to the fact that their values, over the considered production batches were always much lower than the measuring instrument limit of quantification (LoQ).

Fitting a PLS model to the data, a model with a cumulative $R2$ of 0.863 and a cumulative $Q2$ of 0.793 was obtained. Considerable high values for both indicators show that the variability on the yield of FP is almost completely explained by the variability on the quality data of the input material for that production step.

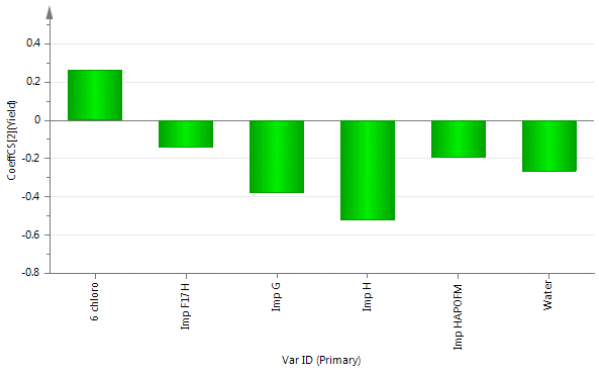

Figure 5: PLS model coefficients for the several impurities present in the intermediary 4 quality data against the yield of FP.

Impurity H appears to be the substance that has the most negative impact on the yield followed by impurity G and the water content. Impurity HAPOFM has a small contribution together with impurity F17H. Impurity 6 chloro has a positive impact on yield. This odd relation is explained in the fact that this substance is present in the production process since the beginning, in the purchased material, and does not purge in any of the production

steps.

Process variables

The quality data of the input material explains most part of the variability in the yield of the final step. However, this fact does not exempt a deep analysis of the process itself as the percentages of variability explained are not additive and so, an also very high value for the way the process is being run can be obtained. As such, process photograph type of data was analysed. This type of data, as explicit in the name, only gives a shot of the process, not the full unfolding of the batch, and is used mainly to give some initial insights on the process and a starting point for process film type of data.

Regression analysis (PLS model) was performed to the data set and although with a very low percentage of variability explained by the model (around 32%), some conclusions can be drawn that were roughly expected with the main one being presented.

For the antisolvent rate of addition, higher values of this variable lead to lower yields. This is expectable as lower rates of addition favor nucleation contrary to higher rates that favor crystal growth [32]. Since the amount of antisolvent loaded is specified in the batch production record and does not vary from batch to batch, a more in-depth analysis was conducted on the antisolvent addition time. This duration is specified as CONFIDENTIAL minutes in the operations manual but there were some batches where it was not followed.

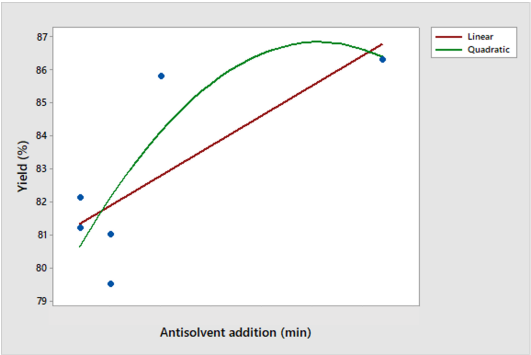

Figure 6: Linear and quadratic regression of yield of the final step against the duration of antisolvent addition on the batches that the indication for CONFIDENTIAL minutes of addition time was not followed. For linear regression, $R2$ equals 0.573 and for quadratic regression $R2$ equals 0.652.

As seen in the graph of figure 6, the higher the time for addition (translated into a smaller addition rate), the higher the yield. The level profile is mandatory to analyse in this case since differences in the addition pattern (which are unknown) are also crucial to the response variable being considered. The improvement action that can be taken out from this analysis is that a level sensor should be integrated into the reactor where FP crystallization takes place.

For the more rigorous and incisive analysis of process film type of data, both the crystallization (and all its successive steps) and the filtration will be the subject of study. Batch Level Modelling (BLM), where the differences in between batches on the variable profiles are modeled against the Y variable and the impact of such differences on the particular problem to be solved is uncovered, was performed.

The BLM model (one component model) for the antisolvent addition step on the crystallization has a $R2$ of 0.6497 and a $Q2$ of 0.459.

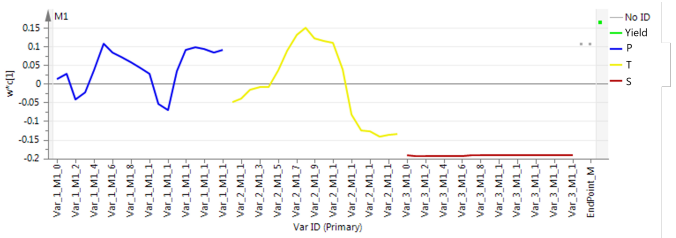

Figure 7: Loadings of the first component given against batch maturity for each variable (pressure as blue, temperature as yellow, and agitator speed as red) for the antisolvent addition step in the crystallization of FP process.

Regarding the temperature impact on the yield, it can be observed that at the beginning of the operation it rises to be strongly positive (meaning that high values of temperature lead to high values of yield) and then decays to be strongly negative. Two improvements could be proposed based on this model: firstly, to promote faster heating at the beginning of the operation since up until the middle of the duration, high temperatures favor high yields and secondly, to lower the final temperature target since the impact of the variable is negative, high temperatures favor low yields. However, the targeted final temperature corresponds to the process reflux temperature, to which the solvent continues to evaporate and therefore cannot be changed.

The final crystallization step BLM model (also a one component model) has the following fitting: $R2$ of 0.648 and $Q2$ of 0.535. A similar amount of variability explained by the model is obtained to the antisolvent addition model. Values in this order of magnitude are considered good and therefore the models being presented can be considered as robust models.

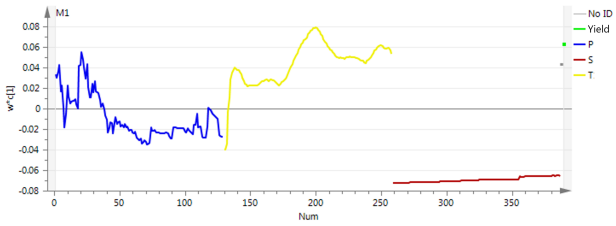

Figure 8: Loadings of the first component given against batch maturity for each variable (pressure as blue, temperature as yellow, and agitator speed as red) for the cooling step in the crystallization of FP process.

Regarding the temperature´s impact on yield, it is clear that it stays strongly positive during the entire cooling operation, apart from the very beginning in which there is an almost vertical ascending contribution. The cooling rate for this operation is specified as CONFIDENTIAL and overall, there are not many differences in the cooling ramp profile from batch to batch. However, what the loadings of the model are exhibiting is that, although in some parts of the operation more evident than others for example the loading peak at x-axis value around 200, lower cooling rates (but still under the process indication of *around* CONFIDENTIAL) lead to higher yields. The contribution of temperature is positive during the final end of the operation, meaning that, higher final temperatures lead to higher yields.

The filtration step of FP process was also analysed. A BLM model with three components was obtained. The cumulative fitting obtained was the following: $R2$ of 0.961 and $Q2$ of 0.768. The Y variable is explained by both the first and second components. However, this does not happen to the third component and so the analysis will be focused only on the two first. The loadings are presented in figure 9.

For temperature, according to the first component there is a positive and strong relationship during all the extent of filtration with yield, meaning that higher temperatures favor yield. That is not the case with the second component, where, although the yield is more poorly explained than by the first component, there is a clear weak relation to the end of the operation with yield. During the filtration, the temperature is not controlled, and the decreasing trend observed for all batches is due to the transfer of the washings, which is done between CONFIDENTIAL and CONFIDENTIAL with no clearer indication. Attending only to the relation shown in the first component an improvement action to be proposed could be to transfer the washings during filtration closer to the upper end of the stipulated process interval.

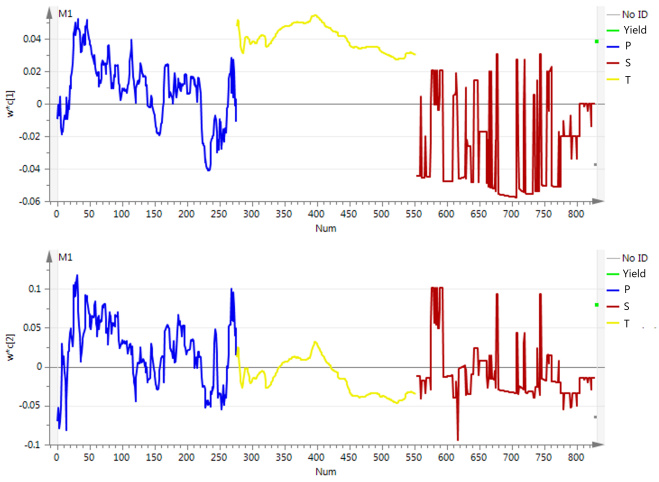

Figure 9: Loadings of the first component (top, $R2$ of 0.497 and $Q2$ of 0.364) and of the second component (bottom, $R2$ of 0.341 and $Q2$ of 0.165) given against batch maturity for each variable (pressure as blue, temperature as yellow, and agitator speed as red) for the filtration step in the FP process.

All the analysis presented had, as response or Y variable, the yield of this production step. Since it is the last chemical step before the final product, it is important to check if the implementation of such measures to optimize yield will damage the process performance regarding quality. This was done by modelling the same operations but considering the final products´s assay as response variable. The findings are that yield and quality go on the same way, *i.e.*, optimizing throughput performance will also lead to an optimization of quality performance.

3.3.2 Intermediary 4 process step analysis

The impurities present in the input material to the FP process step explain most part of the variability in the yield which is the primary problem to be solved on this project. According to the PLS analysis conducted, impurity H has the strongest negative impact on yield followed by impurity G. However, in order to establish concrete improvement actions, a further step back has to be taken in order to evaluate which sections of the intermediary 4 process step are causing the increase in the identified impurities that lead to a decrease in yield.

Process variables

For the subsequent analysis only 8 production batches were considered. The process photograph data for intermediary 4 process step was modeled agains impurity H. A two component model was obtained with the following cumulative fitting: $R2$ of 0.981 and $Q2$ of 0.826. The model coefficients are presented below in figure 10.

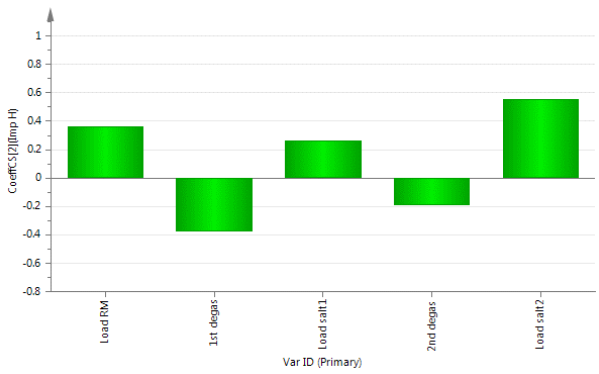

Figure 10: PLS model coefficients for the process variables (process photograph type of data) against impurity H. "Load RM" is the time took for the charging of the raw material (intermediary 3); "1st degas" is the time took on the first degassing step; "Load salt1" is the time took to charge the first inorganic salt; "2nd degas" is the time took on the second degassing step and "Load salt2" is the time took to charge the second inorganic salt.

Impurity H is formed whenever an oxidizing agent is present and so, the process step being analysed has multiple degassing steps prior to the main reaction. The most relevant terms included in the model are precisely the duration of operations deeply related to the entrance or exiting of oxygen in the process vessel, the degassing steps, and the loading of the reactants: a positive relation of the time took to load the reactants is observed meaning that more time spent on loading the reactants yields more impurity H content in intermediary 4 and a negative relation of the degassing duration is observed meaning that less time spent on this operations, more impurity H content will be detected. All these findings are natural since the major motif for the impurity's formation is the presence of an oxidizing agent, in this case, oxygen.

For impurity G, no set of meaningful terms was found to be statistically relevant so, no analysis of the process photograph type of data was conducted considering impurity G as the response variable. However, the process film type of data will be analysed for both impurities, where data was only available for 6 production batches.

The charge of the first salt was modeled against impurity H and the following model (two components), with the cumulative indicators being presented, was obtained: $R2$ of 0.984 and $Q2$ of 0.887. Given the very low relevance of the second component to the current response variable, only the loadings of the first component will be analysed versus batch maturity.

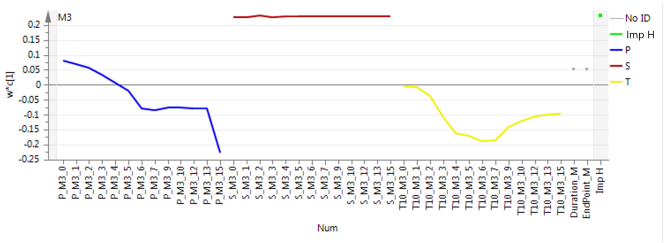

Figure 11: Loadings of the first component given against batch maturity for each variable (pressure as blue, temperature as yellow, and agitator speed as red) for the charge of the first salt during the reaction step of intermediary 4 process.

A strong positive relationship of the agitator speed with impurity H during the entire operation is verified: high values of agitator speed, during the entire operation, yield high values of impurity H in intermediary 4. In regards to temperature, a negative relation is verified, which is especially relevant (more negative) in the middle of the charge but never actually gets positive. The temperature is kept within the process limits (CONFIDENTIAL) but there is the indication of staying as close as possible to the lower limit of the interval. Given the relation presented above, lower values for temperature during the entire operation will increase the content in impurity H thus decreasing the yield in FP process and so the indication to stay at the lower end of the process interval should be changed in order to target the higher end. The temperature relationship with impurity H is maintained in the loading of the second inorganic salt.

Finally, the reaction itself is modeled. A model with two components and a good fitting was obtained: $R2$ of 0.952 and $Q2$ of 0.757. Once again, the second component is much less relevant to the response variable than the first component, allowing it to analyse, on a timely basis, only the loadings for the first component, as they are presented in figure 12.

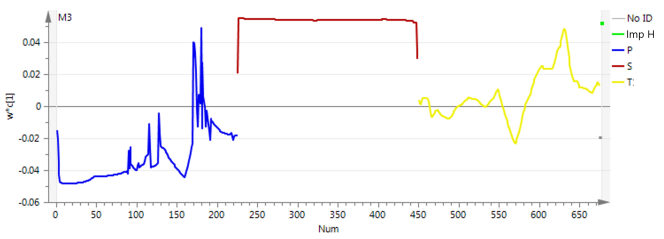

Figure 12: Loadings of the first component ($R2$ of 0.805 and $Q2$ of 0.561) given against batch maturity for each variable (pressure as blue, temperature as yellow and agitator speed as red) for the reaction step of intermediary 4 process.

A closer and combined look at the pressure profile and the loadings for pressure reveals that the peaks in the variable profile coincide with the peaks in the loadings. Setting aside the positive peaks on the loadings, a clear negative and strong relation, especially at the beginning of the addition of the main reactant is observed between the variable and impurity H. In order to decrease impurity H content in the process output, higher pressures should be kept during the initial phase of addition. The main reactant load is done through an auxiliary gas cylinder which is heated in order to keep its content in the gaseous state. A straightforward way to increase pressure in the reactant addition would be to increase the temperature of the gas cylinder. However, this measure could attach countless safety-related additional risks that were not studied. The implementation of this improvement action was left for the production team to carefully analyse.

The analysis considering impurity G as the response variable yielded the same conclusions as the previous models, for impurity H and therefore are not presented.

3.4. Improve

After the statistical analysis results translation into concrete improvement actions, a prioritization method is imperative to apply in order to sort which actions will be tackled first and the number of resources needed. This classification is based on two parameters: impact and effort. The classification based on the former is empirical, based on process knowledge that the production team has been gathering through the batches, and is done on a scale of 1 to 10. The classification based on the latter is performed according to statistical parameters ($R2$ and $Q2$).

The improvement actions that fall in the second quadrant of the matrix (high impact and low effort) can be coined "Quick Wins" and are the ones that should be tackled first because they yield the best return based on the effort. The actions placed on this quadrant were related to the reduction of impurity H content in intermediary 4 through the following of the indication to stay at CONFIDENTIAL (for agitator speed) during all pre-reaction and reaction steps. Considering the increase in yield of FP, two more actions are placed on the second quadrant: the action related to faster heating up until the middle of antisolvent addition, and the action related to the transfer of the washings (during filtration) closer to the higher end of the process temperature interval.

3.5. Control

The final phase of the DMAIC process improvement cycle has one major goal: to sustain the improvements. There are several ways of achieving this goal, however, in this project, only control charts and a summary flowchart will be used.

3.5.1 Flowchart

After an extensive analysis of the primary problem (high variability on the FP process yield) with the root causes being uncovered, it is necessary to implement frameworks within the organization to make sure that the process understanding gained will not be lost. In this line, an internal KPI value for the FP process yield will be established by the production team. Whenever a FP batch has a lower than the KPI yield, an internal investigation is triggered. In order to systematize all the knowledge created during the project and to serve as an aiding tool for the internal investigations, a flowchart was elaborated.

3.5.2 Control Charts

One of the most common tools to apply in the final phase of DMAIC cycles are the control charts that are mainly used for process monitoring. The flowchart for the internal investigations is a great tool for guiding the process of analysing concluded batches and spot what was the cause for a deviation that has already happened. Instead of acting on the problem, a preventive tool like the control charts is essential to actively monitor the process and sustain the improvements in real time. In this way, an interactive Excel file was created, to be filled by the operators with quick and easy to obtain process data (like temperatures displayed and starting and ending times for the operations).

4. Conclusions

Framed on Hovione´s Sete Casas production site continuous improvement plan, the present work aimed at the yield optimization of a production process of a generic corticosteroid API, fluticasone propionate. Six Sigma´s DMAIC process improvement and the problem-solving cycle was the chosen methodology to approach the problem. The project was divided into five separated and clearly defined phases: define; measure; analyse; improve and control, which have proven to succeed on continuous improvement projects. Combined with the DMAIC cycle, a backwards methodology was applied: after identifying the problem in the process output (the high variability on FP process yield), the root causes will be uncovered going in reverse through the API production train.

Ranging from July 2018 to January 2021, 40 production batches were included in the analysis. Over this period, yield took an average value of 81.83% with a standard deviation of 2.22%. The minimum and maximum values were 77.30% and 86.30% which is 11% of the variable mean. High

yield variability leads to biased and uncertain production planning (especially in a multi-purpose installation as is the case) and poor use of company resources since the batch costs do not change with the throughput obtained.

Only the two last steps on the API production train were statistically analysed, the step leading to FP and to intermediary 4. From the models created, several improvement actions were drawn and an Impact Vs. Effort matrix was built. The actions with low effort and high impact, termed "Quick Wins" are related to the agitator speed during the pre-reaction and reaction step regarding impurity H as response variable; to transfer the washings, during FP filtration, closer to the higher end of the process temperature interval and to promote faster heating in the first half of antisolvent addition, during the FP crystallization both regarding yield as response variables.

An interval KPI minimum yield value will be established by the production team and whenever a batch performs under the established KPI, an internal investigation will take place, based on the process understanding generated.

For future work, in order to increase the robustness of intermediary 4 models, more production batches should be incorporated since data was only available for 6. Taking a more holistic approach, this project should serve as the basis for a demystification of MVDA applied to chemical synthesis pharmaceutical processes. These processes are very complex and the black-box statistical approach is the one to take for process improvement.

References

[1] I. C. H. Technical Requirements for Registration of Pharmaceuticals for Human Use. *ICH Q8(R2) - Pharmaceutical Development*. 2009.

[2] J Woodcock. The Concept of Pharmaceutical Quality. *American Pharmaceutical Review*, 2004.

[3] L. X. Yu. Pharmaceutical quality by design: Product and process development, understanding, and control. *Pharmaceutical Research*, 25(4), 2008.

[4] U.S Food and Drug Administration. *Pharmaceutical cGMPs for the 21st Century - a risk-based approach*. 2004.

[5] L. Zhang. Application of quality by design in the current drug development. *Asian Journal of Pharmaceutical Sciences*, 2016.

[6] A. Ferreira, J. C. Menezes, and M. Tobyn. *Multivariate Analysis in the Pharmaceutical Industry*. Elsevier, 2018.

[7] L. Abboud and S. HensleyStaff. New Prescription For Drug Makers: Update the Plants. *The Wall Street Journal*. Consulted on 23/04/2021.

[8] L. X Yu. Understanding Pharmaceutical Quality by Design. *The AAPS Journal*, 16(4), 2014.

[9] D. J. Am Ende and M. T. Am Ende. *Chemical Engineering in the Pharmaceutical Industry: Active Pharmaceutical Ingredients*. Wiley, 2019.

[10] J. Woodcock. Reliable drug quality: An unresolved problem. *PDA J. of Pharmaceutical Science and Technology*, 66(3), 2012.

[11] J. A. Garza-Reyes. Lean readiness – the case of the European pharmaceutical manufacturing industry. *International Journal of Productivity and Performance Management*, 67(1):20–44, 2018.

[12] B. Chatterjee. *Applying Lean Six Sigma in the Pharmaceutical Industry*. Routledge, 2014.

[13] D. Bellm. Operational excellence in the pharmaceutical industry – an architecture for emerging markets. Master's thesis, University of St.Gallen, 2015.

[14] H. Gebauer. Lean management practices in the pharmaceutical industry. *International Journal of Services and Operations Management*, 5(4), 2009.

[15] T. Friedli. Pharmaceutical OPEX - The Next Generation. *Life Science Leader Magazine*, 2013.

[16] T. Friedli. *Leading Pharmaceutical Operational Excellence*. Springer, 2013.

[17] A. P. Ferreira. Multivariate analysis in the pharmaceutical industry: Enabling process understanding and improvement in the PAT and QbD era. *Pharmaceutical Development and Technology*, 20(5), 2015.

[18] L. X. Yu. The future of pharmaceutical quality and the path to get there. *International Journal of Pharmaceutics*, 528(1-2), 2017.

[19] S. Nussey. *Endocrinology: An Integrated Approach*. BIOS Scientific, 1st edition, 2001.

[20] W. Ericson-Neilsen. Steroids : Pharmacology , Complications , and Practice Delivery Issues. *Oschner Journal*, 14(2), 2014.

[21] H. Schacke. Mechanisms involved in the side effects of glucocorticoids. *Pharmocology and Therapeutics*, 96, 2002.

[22] K. Remien. *StatPearls: Fluticasone*. 2021.

[23] Fluticasone Propionate Cream WebMD. https://www.webmd.com/drugs/2/drug-8786/fluticasone-propionate-topical/details. Consulted on 01/04/2021.

[24] Fluticasone Propionate Nasal WebMD. https://www.webmd.com/drugs/2/drug-77986-245/fluticasone-propionate-nasal/fluticasone-spray-nasal/details. Consulted on 01/04/2021.

[25] K.S. Krishnamoorthi. *A First Course in Quality Engineering: Integrating Statistical and Management Methods of Quality*. CRC, 2011.

[26] T. M. Kubiak. *The Certified Six Sigma Blackbelt Handbook*. ASQ Quality Press, 2009.

[27] P. S. Pande. *The Six Sigma Way: How GE, Motorola, and Other Top Companies are Honing Their Performance*. McGraw-Hill, 2000.

[28] H. C. Ott. Six Sigma and Lean: Quantitative Tools for Quality and Productivity - Yellow Belt Certification handouts. *TUM - edX Platform*, 2020.

[29] G. P. Agre. The Concept of Problem. *A Journal of the American Educational Studies Association*, 1982.

[30] F. Voehl. *The Lean Six Sigma Black Belt Handbook: Tools and Methods for Process Acceleration*. 2013.

[31] S. Estrada-Flores. Development and validation of "grey-box" models for refrigeration applications: A review of key concepts. *International Journal of Refrigeration*, 29(6), 2006.

[32] H. H. Tung. *Crystallization of Organic Compounds: An Industrial Perspective*. Wiley, 2008.