Simulation Model of Quality Control Laboratory in Pharmaceutical Industry

Andrea Costigliola
andrea.costigliola@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

2 November 2016

Abstract

In the last decades the pharmaceutical industry has been through many changes, as a result of increased competition, patents expiration and increased drug quality standards requirements. In order to compete against peers, drug manufacturer started to be more concerned on achieving operational excellence through the optimization of all process involved in drug manufacturing. During its development, a drug must be constantly monitored with laboratory tests. In this scenario, quality control laboratories are critical components in drug manufacturing, and inefficiencies in laboratory management can have a major impact on the overall supply chain service level. The aim of this thesis is to build a Discrete Event Simulation model of a Quality Control laboratory. To achieve this objective, a generic framework for information treatment and organization was built. In particular, information coming from different databases was organized into a single one that was used as input to a discrete event simulation model. The proposed model represents in detail the work flow of a quality control laboratory and it is intended as a support tool for planning, scheduling and decision making. The model was validated using real data, and it resulted effective to estimate performance parameters such as, system throughput, equipment usage rate, system responsiveness and tasks processing times. Furthermore, the simulation model was tested with an alternative scheduling policy to evaluate how modifications on the system may improve its performance.

Keywords: Quality Control, Discrete Event Systems, Discrete Event Simulation, Scheduling, Pharmaceutical industry

1. Introduction

In the last decades drug manufacture evolved driven by external economic forces, patents expiration and increased competition. In order to maintain their competitive advantage, pharmaceutical companies organized themselves into complex organizations (supply chain and contract manufacturing networks) and started to be more concerned on achieving operational excellence through the optimization of all the processes involved in drug manufacturing: supply of raw materials, logistic operations, chemical processes modeling, management of quality control laboratories, etc.

Pharmaceutical companies operates in one of the most competitive and regulated markets, and compliance with Good Manufacturing Practices (GMP) and Good Laboratory Practices (GLP) is mandatory for marketing any drug. Those regulations were introduced to ensure that every pharmaceutical product meets safety and quality requirements in a systematic fashion.

The product life cycle must be constantly monitored with laboratory tests to ensure that it meets the intended quality and purity characteristics. In this scenario, quality control laboratories play an important role in the drug manufacturing process. Laboratories are responsible for quality, safety and efficacy of new medicines, and their management is a complex task that involves resource planning and scheduling, analysis prioritization, results documentation, etc.

Quality control laboratories are critical components in drug manufacturing, and inefficiencies at laboratory level may delay obtaining results, affect negatively their quality and can have a major impact on the overall supply chain service level. This situation can be magnified in the case of a contract manufacturing organization that produces goods under the brand of its clients, and therefore has to deal with a large number of projects. In pharmaceutical industry, scheduling is often done by supervisors experience. Laboratory information management systems (LIMS) used by pharmaceutical companies only track the analyses performed and lack of some features (i.e. information on processing times, work flow) essential for planning, scheduling

and stock management.

From a research perspective, studies regarding the concrete production industry have been often limited to the manufacturing processes. Advances in informatics, data analysis and knowledge management made industries aware about the power of information. This information can be organized and generate knowledge in order to improve the quality of the services and manufacturing processes. As a result, informatics is being always more incorporated in the industrial setting, and nowadays *industrial informatics* represents an important field of study, and informatics is no more related only to the Information Technology (IT) services and infrastructures, but also to the design, simulation and modeling of manufacturing processes.

The objective of this work is to build a generic framework for modeling and simulating a quality control laboratory of a contract manufacturing organization. In particular, this work uses the concepts of Discrete Event Systems theory for modeling purposes. A model of the system is simulated using Simio simulation software, and the analysis of the output will serve as basis for planning, scheduling and decision making in quality control laboratories.

2. Quality Control in the Pharmaceutical Industry

In order to achieve quality in a manufactured product, three main tasks are needed [7]:

- Quality planning;
- Quality control;
- Quality improvement.

Planning refers to the identification of quality goals, customers needs and to the development of specific manufacturing and control processes to meet the requested quality standards. Control refers to the evaluation of the manufacturing process performances towards quality goals. Quality improvement is the process that creates new needs and procedures for higher quality standards.

The process that lead from the discovery of a new promising compound to its marketing, also referred to as cycle time, passes through four main development phases and may take many years. In this period, a continuous and recursive process of quality planning-control-improvement is needed. Analytical chemistry is a measurement science which develops and uses methods in order to perform quantitative and qualitative analysis. Given a certain compound, a specific analytical method has to be designed and validated to be used in quality control. A method should be developed with the goal to rapidly test preclinical samples, formulation prototypes and commercial samples [3].

Although analytical methods development and validation are fundamentals for any drug research and development program, there are other operations directly related to the manufacturing process that must be performed: *In-Process control* and *stability tests*. In-Process control are tests performed during the manufacturing process to monitor the process and provide insights for its on-line correction when necessary. Stability tests are performed on active pharmaceutical ingredients (APIs) and final pharmaceutical products (FPPs) to recover information on how the quality of the drug varies with time under the influence of external factors, such as temperature, humidity and light exposure.

2.1. Quality Control laboratory management

Due to the challenging scenario described, the optimization of the operations involved in the process of drug manufacturing is nowadays a necessity. Optimization methods can not only improve the quality but also minimize process operational costs and reduce variation in the product [13]. Products quality and operational efficiency depend on timely delivery of quality analysis results.

Laboratory management is a complex task which includes resources (personnel and equipment) scheduling, analysis prioritization, results evaluation and documentation. Inaccurate scheduling policies may delay obtaining results, affect negatively their quality and may generate financial losses because of drug product degradation. Quality control efficiency can be essentially improved by analysis rational scheduling [5].

Planning is a process used to evaluate how many resources and what resources are needed to perform a job. In a lab environment, planning evaluates the correct number of analysts and instrumentations needed to perform quality control operations. Scheduling refers to the allocation of jobs to available resources and its prioritization. *Scheduling* defines what job should be done first and by who [12].

Operations research (OR) provides methods and tools for the optimization of similar planning and scheduling problems [5]. As suggested by Maslaton [10], resource analysis should be done at regular intervals (i.e. monthly, yearly), based on the dynamics of the forecast. The author proposes the development of a generic framework to organize laboratory information to be used for laboratory workload forecasting.

In general, in a quality control laboratory many samples concur for the same resources (i.e. analysts and machines). An optimal schedule will ensure assigning the samples/tests to the best available resource and that the right tests are started at the right time with no delays [11].

An interesting solution for laboratory personnel

scheduling is the one proposed by Boyd et al. [2], where the authors applied a genetic algorithm for analysts scheduling. Their program maximizes the value of a fitness function that measures how well a given work shift scheduling of analysts and their skills matches a set of tasks.

Dudnikov et al. [5] developed in 2011 a laboratory analysis planning system to schedule analyses on the available resources. They considered two types of samples for their system: periodic samples and unplanned samples. They created a ranking in order to prioritize the analyses and prepared a 24 hours scheduling considering only the periodic samples.

Literature is short of examples of generic optimization frameworks applicable to pharmaceuticals quality control management. However, Schäfer proposes a set of terms and definitions that may be used as basis to build a consistent framework for laboratory management [14]. The author proposes the modeling of all components interacting on the workbench (i.e samples, instruments, sensors, results, information systems) and provides a description of a schematic scheduling process that can be applied in quality control laboratories.

3. Discrete Event Simulation

A *simulation* can be defined as the imitation of a real world process or system over time. Simulation involves the generation of an artificial history of a system (i.e. process, facility) and its observation in order to get insights on system operating conditions [1]. Simulation has been widely used over the last 50 years and it is nowadays one of the most used operations research (OR) and management-science techniques, if *not* the most widely used [8].

Following the definition proposed by Schmidt and Taylor in 1970, a system can be defined as a collection of entities that interacts towards the accomplishment of some goal [15].

Discrete event simulation models are stochastic and the dynamic of the system is driven by events happening at discrete points in time [16]. In *simulation* a system model is evaluated numerically with a computer, and relevant information is collected to *estimate* the behavior of the system [8]. The availability of special-purpose simulation languages made possible the large diffusion of computer simulation methods. Nowadays, simulation represents an economic and flexible choice that allows to recover information on a system, simulate changing scenarios and get insights on its dynamics.

3.1. Discrete Event Systems.

A *discrete event system* is a particular type of dynamic system whose state space is naturally described by a discrete set, and state transition are observed at discrete points in time. Those points in time are associated with "events" [4]. An event can be seen as "something that happens" and can cause a transition from one state value to another.

Definition 1. A discrete event system (DES) is characterized by:

- Set of possible events: $E = \{e_1, e_2, ..., e_n\}$;
- Discrete state space: X;
- Event driven dynamics: $x_{k+1} = x(\delta_k, e_k), k \in \mathbb{N}$, where:

  x_{k+1}: next state after the k_{th} event happens;

  e_k: k_{th} event happened from the considered point in time;

  $\delta : X \times E \mapsto X$: state transition function.

From the definitions above, it is clear that the behavior of a system depends on events that drive the system into a particular state. During the study of a discrete event system it is possible to consider different types of event sequences: *logic event sequences* and *timed even sequences*. In logic sequences the time when an event occurs is not explicit. In timed sequence, the time of occurrence of an event is explicit and it can be deterministic or stochastic. In a simulation study we are interested both in the logical and dynamics behavior of a system and thus, timed event sequences are a more appropriated modeling tool. In this context, timed Petri nets have been used to describe a model of the system.

3.2. Timed Petri net

Petri nets were first developed by Carl Adam Petri in the early 1960s. They provide an explicit representation of the transition function of a system. A Petri net is a device that manipulates events according to predefined rules. A system modeled as a Petri net can be easily described in graph form, resulting in a *Petri net graph*, that is very intuitive and self-explanatory [4].

A Petri net is a *bipartite graph* with two type of vertexes: *places* and *transitions*. In a Petri net events are associated with transitions. Places contain information on the state of the Petri net. In order for a transition to "fire" (occur), the Petri net must meet particular conditions.

Definition 2. A *Petri net graph* is a 4-tuple (P, T, A, w), where:

P is the finite set of *places*

T is the finite set of *transitions*

$A \subseteq (P \times T) \cup (T \times P)$ is the set of arcs from places to transitions and from transitions to places

$w : A \mapsto 1, 2, 3, ...$ is the weight function on the arc set.

To describe the evolution of a Petri net it is necessary to define the concept of *token*. A *token* is an entity that occupies a place in a Petri net. The way tokens are assigned to a Petri net graph defines a *marking*, that represents explicitly the state of a Petri net graph and defines the condition under which an event (transition) "is enabled" and can fire.

Definition 3. A *marked* Petri net is a five-tuple $(P, T, A, w, \mathbf{m})$ where (P, T, A, w) is a Petri net graph and $\mathbf{m}$ is a $P \times 1$ vector containing the number of tokens in each place of the net:

$$\mathbf{m}^T = [m(p_1), m(p_2), ..., m(p_n)] \in \mathbb{N}^n.$$

A transition is enabled and can fire if the number of tokens in the places upstream the transition, contain a number of tokens greater or equal than the weight of the arcs that links those places to the transition. When a transition fires, it leads the system to a new state.

In discrete event simulation we are interested to the system evolution in time and, with this purpose, it is necessary to define a *clock structure* associated with a set of transitions $T_D \in T$. Transitions in this set are called *timed transitions*. When a timed transition is enabled, it does not fire immediately, but after a time delay given by the clock. In this case, the clock regulates the happening of the events, and thus the firing of timed transitions in the Petri net.

Definition 4. The *clock structure* associated with timed transitions of a marked Petri net is a set:

$$V = v_i : t_i \in T_D$$

of *time sequences* $v_i = \{v_{i,1}, v_{i,2}, ...\}$,

$$t_i \in T_D, \quad v_{i,k} \in \mathbb{R}^+, \quad k = 1, 2, ...$$

It is now possible to define a timed Petri net, that will be used to model the discrete event simulation system.

Definition 5. A *Timed Petri Net* is a six-tuple $(P, T, A, w, \mathbf{m}, \mathbf{V})$, where $(P, T, A, w, \mathbf{m})$ is a marked Petri net, and $\mathbf{V}$ is a clock structure associated with transitions $t_i \in T_D$.

In a timed Petri net graph, timed transitions are labeled using the notation v_i, while zero delay transitions are labeled keeping the notation presented for Petri net graphs (t_i) (see Figure 1).

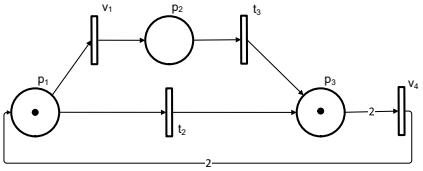

Figure 1: A timed Petri net with two timed transitions.

4. Simulation model of a Quality Control laboratory

The main objectives of this project are to gather knowledge on the *Quality Control system* of a contract manufacturing organization, build a discrete event system model and perform simulations to mimic its behavior. The first part of the project has been dedicated to the study of the quality control laboratory and to the identification of the objectives of the study. In particular, all the stakeholders were interested in estimating *equipment and labor utilization*, and the responsiveness of the laboratory (i.e. the time needed to perform an analysis and the time it takes with the existing configuration). From an operational perspective, different aspects have been addressed during the study of the real system: available resources (equipment, laboratory instruments, personnel) and their characteristics/specificities, inputs to the system (jobs, activities), and outputs (results of an executed job/activity).

The study of a dynamic system should be performed in a systemic way. It is important to understand how products and information flow through the system, and to identify all the processes and variables of the real-world system. According with the Object-Oriented programming paradigm provided by modern simulation softwares, the system can be modeled as composed by different objects and entities that interact to provide the desired output. Following this paradigm, the main processes have been modeled as independent systems with their own dynamics, logics and properties, and have been later added to a hierarchical model in order to build the entire system.

4.1. System Description

A quality control laboratory performs a series of tests on a sample using analytical techniques and methods. Usually we identify an *analytical technique* with a specific equipment, while the term *analytical method* is sample specific and describes how the analysis must be performed. For this study, six analytical techniques have been considered critical to the laboratory efficiency and the relative instrumentations have been modeled. Those techniques are: *High Performance Liquid Chromatography* (HPLC), *Gas Chromatography*, *Particle Size Analysis* (PSA), *Karl Fischer Tritration* (KF), *Dif-*

ferential Scanning Calorimetry (DSC) and *X-Ray Powder Diffraction* (XRPD) [6].

Even if those techniques are very different from each other, and so are the relative equipment, it is possible to identify steps that are common to all those analyses:

- System preparation;
- Equipment setup;
- System suitability;
- Sample preparation;
- Analysis;
- Analysis of the results.

Some of those steps are performed exclusively by the analyst or by the equipment, while others are performed on the equipment in presence of the analyst. When a sample is received in the laboratory and it is scheduled to be analyzed, the analyst starts the analysis preparing samples and additional solutions/materials necessary to execute the system verification. The next step is the setup of the equipment performed. Once the equipment is ready, it is needed a verification (*System Suitability*), that must be performed before the test, to ensure that the equipment is adequate for a particular analysis. This step is performed by the equipment autonomously, and the analyst has the task to verify several parameters at the end of the *System Suitability* run. Once the system is verified, the equipment is ready to analyze the sample. At the end of the run the analyst disassembles the equipment (unless there are new samples to be analyzed with the same method), and analyze the data of the analysis.

Analysts and equipment can be considered as the processing units or resources of the system. Jobs and activities are defined by the sample and by the analytical methods that have to be used to analyze the sample. Thus, samples can be considered as the input of the system. In the simulation software paradigm they can be defined as entities that flow through the system and need to be processed in a given number of processing units according to the analytical techniques and methods to be used.

4.2. Data Processing

One of the biggest challenges of this project was gathering and processing information. In fact, information was distributed on different databases, and in some cases it was stored in a document repository in text form. For the purpose of this project it was necessary to extract the relevant information from all these data sources and merge it into a unique database to be used for modeling. In the section above, the main components of the system have been identified: analysis work-flow, entities (samples) and processing units (analysts and equipment). The next step in the modeling process is to estimate *entities arrival rate*, *process logics* and *processing times*.

This information can be retrieved from different data sources. In particular the Laboratory Information Management Systems (LIMS) contains information on the analyses performed and the number of samples entering the quality control laboratory. Another source of information that is very powerful even if not structured is the *document management system* (DMS). In this database all the document with the description of the analytical methods are stored. Moreover, to estimate processing times, a time study was performed to gather information and have a better characterization of the processing tasks.

LIMS data The Laboratory Information Management System (LIMS). A LIMS provides mechanisms to build automated tasks and to integrated them [9]. It is also a powerful tool used to track, document and review every analysis performed in a laboratory. LIMS is not only a tool that provides support to the analytical chemistry area, but in times of big data analysis, knowledge work automation and in general knowledge management it represent an opportunity to improve the quality of the analytical operations. In this context, LIMS database has been used to recover information on every analysis. In particular, each entry of the data table contains information on sample number, name of the project, analytical method used, equipment used and analyst that performed the test. Moreover, it has informations about the total number of analyses performed.

LIMS database can be considered as the central source of information. It is not only the most complete database, but data is structured and easy to access.

Information extraction algorithm An information extraction algorithm has been developed to analyze chromatography analytical methods report. An analytical method report contains information about: run time, analysis type, solutions to be prepared and number of injections. It was then implemented an information extraction algorithm using the *R* programming language, to identify and extract information. The algorithm converts a *pdf* document into a string array variable and finds table structures from a *docx* file of the same document storing them into a *data frames*. It then identifies the language of the document and choses the relative vocabulary to be used to match words and

patterns. The algorithm extracts the following information:

- Analytical method name
- Number of samples to be prepared to run the system suitability;
- Number of samples to be prepared to run an analysis;
- System suitability run time;
- Analysis run time.

Time study and work measurement In a quality control laboratory, analysts are responsible for sample preparation, system setup, equipment cleaning and analysis data processing. To estimate those parameters a *time study* has been designed to gather information on tasks processing times and to divide equipment "hands-on" and "hands-off" tasks. With this purpose, it was asked to analysts to collect during a period of one moth by filling a manual form model.

Complete Database The couple sample/analysis has been identified as the central figure of the quality control laboratory system. In fact, a sample should carry on all the information necessary to analyze it. Thus, it seemed appropriate to organize data by the unique key given by the couple sample/analysis. All the collected data were then organized into a unique database, whose UML representation can be seen in figure 2. LIMS has

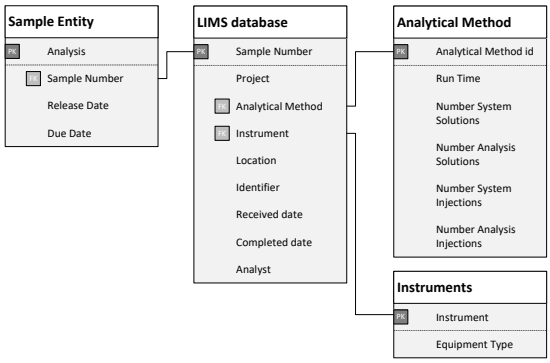

Figure 2: UML representation of the designed database.

been considered as the central information source in the design of the database structure. Other tables, such as *Analytical Methods* and *Instruments* have been created using respectively the output of the information extraction algorithm and by grouping the list of instruments per analytical technique. *Sample Entity* table condenses all the information needed to process a sample and it could be effectively used for planning and scheduling.

4.3. Modeling

The developed simulation model of the quality control laboratories is composed by the following objects:

- Equipments;
- Analysts;
- Work locations;
- Samples.

An equipment can be seen as a series of processes needed to perform an analysis. Those processes are: sample preparation, system suitability, analysis preparation, analysis run and data processing. Even if not all those tasks are performed at the equipment location, the modeling strategy chosen is process oriented and the spatial components have been added to those processes selecting the physical place in the quality control laboratory model where the analyst has to go to perform the requested task.

An analyst has been modeled as a secondary resource of the system, which means that his/her presence is necessary to perform specific tasks. Analysts work in shifts and their availability in the laboratory model varies according to the work shifts.

Samples have been modeled as input entities that travel in the system to perform certain tasks. Each sample has properties that define the type of equipment needed for processing it, and informations about run time and number of preparations. A sample can be processed on an equipment without the need to perform the system suitability if the last sample processed on the same equipment used the same analytical method. In particular, it has been assumed a validity of 24 hour for a system suitability.

A generic machine model has been implemented to simulate the analysis process. A timed Petri net have been used to represent the underlying discrete event system (figure 3). The Petri net is characterized by a set of places P, that represents the five tasks described above, more auxiliary tasks needed to avoid situation like performing the system suitability and the analysis at same time. It has four timed transition corresponding to processing times for each task.

When samples enter the system they wait in a queue and are scheduled according to a policy that estimates the lateness of the sample and gives higher priority to samples with higher lateness. Lateness L_i is defined as the difference between the expected completion time C_i and the due date D_i:

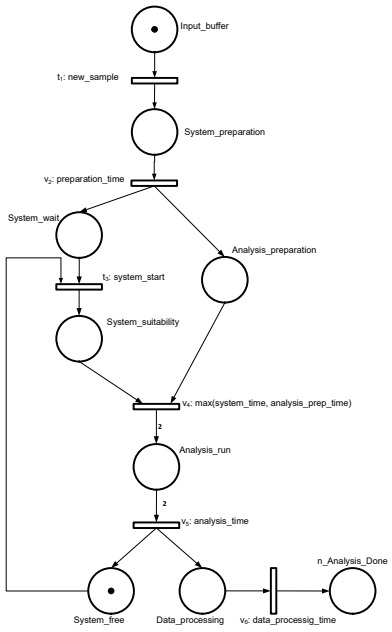

Figure 3: Discrete Event System model of a generic equipment.

$L_i = C_i - D_i$. This scheduling policy tends to minimize the maximum lateness. With this approach, the completion time of an analysis in the simulation model should be closer to the one existing in the real data. .

4.4. Processing time distributions

Many of the processing times in the model are stochastic in nature, while others are deterministic. Deterministic processing times are mostly associated with equipment tasks. However, part of those processing times are assumed to be stochastic given the empirical nature of the available data. Processing times for HPLC and GC equipments are assumed to be deterministic, since they are method specific and exact information was retrieved using the *information extraction algorithm*. For other equipment types and for tasks depending on the work of an analyst (i.e. equipment setup, sample preparation, data processing), the processing times have been obtained fitting curves to the empirical data. Given the limited data available for estimating the processing times, uniform and triangular distributions have been used to approximate the data.

4.5. Simio Model

The described model has been implemented in Simio in order to run simulation with the model. In Simio it was created a library containing the following objects:

- **Equipment**: that is the implementation of the generic machine model. This object has been used to build complex objects, such as machine groups.
- **HPLCs**: equipments group containing all the HPLC considered in the study. In this model the physical location of the equipments and the lab benches have been implemented.
- **GCs**: equipments group formed by the Gas chromatography equipments.
- **KFs**: as for the GC and HPLC equipments a machine group of the Karl Fischer titration equipments have been implemented as an object.

The Equipment object represents a generic implementation applicable to all types of machines. Moreover, processing times, secondary resources needed for processing (i.e. analysts) and physical locations of bench, computer and equipment are set as inputs to the model and can be easily modified externally. The developed object library has been used in a general model that includes all the equipment existing in the system

The simulation model was developed under the following assumptions:

1. *The couple sample/analysis has been considered as a single entity.* Every sample entity is processed using only one machine;
2. *The system suitability has a validity of 24 hours.* This means that the system suitability does not need to be re-executed within 24 hours, unless a new sample using a different analytical method is scheduled to the same machine;
3. *Analysts can perform every type of work regardless the work shift they are allocated to.* In the real system every work shift is allocated only to few operation types (i.e. in-process control, stability, development).

Under those assumptions, the implemented model imitates the original system by driving each sample towards the equipment that was used in the real system to process that particular sample. With this configuration it is possible to estimate which equipments have been used the most and also it is possible to compare the throughput of the simulation model with the real system.

For simulation purposes, a modified model have been implemented in order to simulate the behavior of the system using different scheduling techniques. In particular, an *ad-hoc* scheduling rule have been created for this second model. This rule tries to allocate a sample to equipments that are

"less loaded", in case equipments that are performing the same method have four or more samples in queue.

5. Simulation Results

The main objectives of a simulation study are to measure the performance of a real system and to evaluate how modifications on the system can affect its performance. In this study, the following aspects were considered: *system throughput*, *tasks processing time*, *equipment usage rate*, *system responsiveness* and *employee utilization*. The first three metrics were compared with the available data to validate the model. In particular, it was considered a one year period data (2015) for validation purposes. Statistical independent replications of the simulation model, using the same input parameters, were run to retrieve information on the behavior of the model.

In this context the throughput is defined as the total number of samples processed by the system over the total number of samples entering the system. The processing time for each task of an analysis can be computed by dividing the total processing time of each process by the number of analysis performed. This time represents an average and it must be complemented with its standard deviation. The equipment usage rate is defined by the formula:

$$Usage_{\%} = \frac{EPT}{TRT},$$

where EPT is the Total Equipment Processing Time, and TRT is the Total Run Time. *System responsiveness* is measured as the average time needed by the system in a steady state to perform an analysis. *Employee utilization* refers to the percentage of time an analyst is busy performing bench work (i.e. preparing samples, processing data) over the total time available.

5.1. Model Validation

To validate the model, several simulations have been performed over a time period of one year (2015). In this experiment the system was able to perform all the analyses that were input to the system on time, resulting in a throughput of 100%.

Moreover, the equipment usage rate generated by the simulation model has been compared with real data. In figure 4 are shown the usage rates for the selected HPLC equipments coming from the simulation model and from the real data. Similar results were found for GC equipment. The difference between real data and simulated data is within 1.5%. Given the lack of real data to use for comparison purposes, it is difficult to validate the model behavior for other classes of equipments.

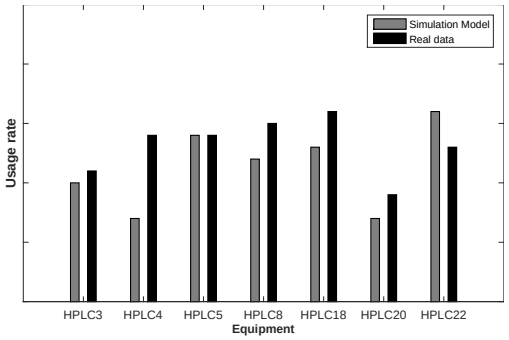

Figure 4: HPLC usage rate comparison between simulation and real data.

5.2. Real system model simulation

To simulate the real system, LIMS data has been used to define the number of analyses performed. The model tries to imitate the real system by replicating the analyses on the same equipment that were used during the considered period.

Throughput For the considered period, the simulation provided a throughput of 100% in all the simulation runs (50 statistical independent runs). The model was able to process all the samples, in the system with no delays on the due date.

Tasks processing time Processing times were estimated using both real data (for chromatographic techniques) and empirical data obtained observing the real system. In figure **??**, percentages of each task for HPLC equipment can be found. In particular, those data agree with the time distributions estimated, and as it can be seen, tasks related with the system suitability represent the largest part of the analysis.

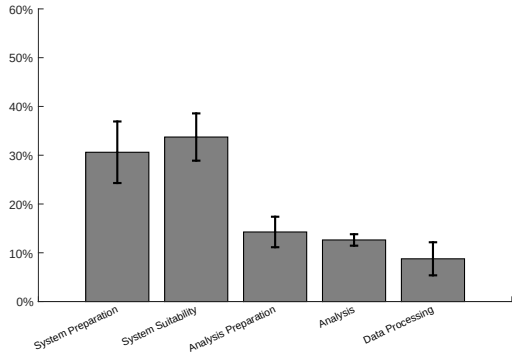

Figure 5: HPLC: tasks percentage on total time.

Equipment Usage rate Equipment usage rate is a metric used to understand how much time the equipment is performing active work. This value is calculated over 24 hours per day, 7 days per week.

However, this is not the real scenario, since those equipments are not able to perform analyses in a complete autonomous way, and they still need an analyst to *start* an analysis. Equipment usage rates have been found to be very small (around 10%). This can be explained by the fact that the number of analyses considered as input to the model was lower than the real number for some of the machines. In fact part of the analyses are not present in LIMS database (i.e. validation analyses, analyses referring to non GMP work).

Employee utilization In quality control laboratory, employees represent an important part of the system. Their work have been calculated considering all the time they perform active bench work, such as a preparation, a setup on a given equipment, data processing, etc. In this model the utilization rate has been estimated around the 35% considering all analysts. This value may seem not very high, but there exists a lot of hidden work that the model is not considering, such as writing analysis records.

5.3. Capacity experiment

To test the system responsiveness the two versions of the implemented model (which differ for the scheduling policy) have been tested over three different scenarios, with the objective of evaluating the capacity of the two configurations in processing samples:

- **Scenario 1**: Model with the original number of samples to process;
- **Scenario 2**: The number of samples entering the system is increased by the 25%;
- **Scenario 3**: The number of samples entering the system is increased by the 50%.

In the first scenario, the modified model is much more responsive than the original system. In fact, for an HPLC analysis the time in system retrieved from the simulation is of 31 ± 14 for the original model and 15 ± 1 for the model with the new scheduling technique. Similar results have been obtained for the Gas Chromatography equipment, where the improvement have been found to be around the 30%. For other types of equipment the situation is similar. However, in Scenario 2 and 3 the second model does not seem to behave better than the first model. In Scenario 2 it is still able to perform faster HPLC samples, but it has worse performances on other equipments. The situation gets even worse in the third scenario, where the second model is very slow in processing analyses and for equipments such as the X-ray powder diffractometer the average time in system for Model 2 is 312 hours against the only 12 hours of Model 1. This situation can be explained by the fact that since the second model tries to allocate more samples to empty equipments, and there are not enough analysts to process them all, part of the samples suffers a huge delay. On the other hand, the first model that uses always a lower number of equipments has to perform less system suitabilities and analysts are often available to end a task.

Comparing the two models and the different scenarios it is clear how scheduling affects a lot the performance of the system. Rational scheduling can improve the productivity of a quality control laboratory, while poor scheduling technique can lead the system to a state in which it is no more able to process jobs on time.

6. Conclusions

The main objective of this project was the development of a simulation model of a quality control laboratory. The model represents in detail the entire work flow of the quality control laboratory and it is intended as a support tool for planning, scheduling and decision making.

For the study of this kind of systems, Discrete Event Simulation was found to be an appropriate tool. Moreover, Petri net formalism provides a graphical way to represent the system that can be used as guideline to implement the model in a simulation software. To the scope of this thesis, the academic version of Simio software was used both for modeling and simulating. In order to make easy modifications to the model, the inputs have been structured in data tables and Simio objects for equipments and groups of machines have been implemented into an external library that can be expanded and used to model other laboratories or to test modification on the proposed laboratory model. Also, the implementation of a generic machine model applicable to different types of equipment make the model flexible to be rearranged in different scenarios.

The model was appropriate in estimating task processing times, number of resources needed, capacity of the laboratory, equipment usage rate, system throughput and responsiveness. Given a well structured input table, it can be used for planning purposes, such as the prediction of the resources needed to meet all the due dates. The performance of the quality control system was evaluated on different levels. The simulation model was able to perform all the analyses meeting the due dates. For what concerns equipment usage rate it was found to be very small (around 10%). The main reason for this, is that data coming from LIMS was incomplete, and as consequence part of the analyses (i.e. validation analyses and non GMP tests) was not considered.

6.1. Future Work

Future works include the improvement of data quality, through a review of the processing times. Moreover, it is necessary to complement LIMS data in order to estimate correctly the number of samples processed by the quality control system. Further on, improvement to the model can be done by reviewing the assumptions made in section 4.5. An additional feature that could be considered in a next version of the model is the stock management. In this project, solvents and solutions have been considered infinite resources, and it would be interesting to add constraints to the model to evaluate limitations coming from lack of stock.

Acknowledgements

This research was supported by *Instituto Superior Técnico* and *Hovione Farmaciência S.A.*. The author would like to thank Prof. João Sousa, Prof. Susana Vieira, and Dr. Filipe Ataíde for the help given during the development of this project.

References

[1] J. Banks, J. S. Carson II, B. L. Nelson, and D. M. Nicol. *Discrete-event System Simulation*. Prentice Hall International, 4th edition, 2004.

[2] J. C. Boyd and J. Savory. Genetic Algorithm for Scheduling of Laboratory Personnel. *Clinical Chemistry*, 47(1):118–123, 2001.

[3] J. Breaux, K. Jones, and P. Boulas. Analytical Methods Development and Validation. *Pharmaceutical Technology*, pages 6–13, 2003.

[4] C. G. Cassandras and S. Lafortune. *Introduction to Discrete Event Systems*. Springer Science & Business Media, 2009.

[5] E. E. Dudnikov, A. Sidorenko, and E. N. Khobotov. Laboratory Analysis Planning System for Industrial Enterprises. *Automation and Remote Control*, 73(11):1910–1915, 2012.

[6] D. Harvey. *Modern analytical chemistry*. McGraw-Hill, 2000.

[7] J. M. Juran and A. B. Godfrey. *Quality handbook*. McGraw-Hill, 1999.

[8] A. M. Law. *Simulation Modeling and Analysis*. McGraw-Hill, 5th edition, 2015.

[9] R. Mahaffey. *LIMS: Applied Information Technology for the Laboratory*. Springer Science & Business Media, 2012.

[10] R. Maslaton. Resource Planning in QC Laboratories. *Pharmaceutical Engineering*, 32(2):1–6, 2012.

[11] R. Maslaton. Resource Scheduling in QC Laboratories. *Pharmaceutical Engineering*, 32(6):1–6, 2012.

[12] M. Pinedo. *Scheduling*. Springer, 5th edition, 2015.

[13] M. A. Rahim and K. S. Al-Sultan. *Optimization in Quality Control*. Springer, 1997.

[14] R. Schäfer. Concepts for Dynamic Scheduling in the Laboratory. *Journal of the Association for Laboratory Automation*, 9(6):382–397, 2004.

[15] J. W. Schmidt and R. E. Taylor. *Simulation and Analysis of Industrial Systems*. Richard D. Irwin, 1970.

[16] A. A. Tako and S. Robinson. Comparing Discrete-Event Simulation and System Dynamics: users' perceptions. *Journal of the Operational Research Society*, 60(3):296–312, 2009.