Data Driven Simulation in the Health and Hospital Field: a Model for the "G. Gaslini" Clinical Pathology Laboratory Workout

Roberto Revetria, Pietro Giribone, Francesca Oliva, Maurizio Schenone

***Abstract*—** The use of IT tools and simulation in the health and hospital field is growing up to improve the efficiency and reduce the costs. The concepts used in the simulation of hospitals are more or less the same of those used in the industrial system's one, but in the first there are more variability connected to the high presence of human intervention. This leads to the use of logic-type models, with the simulator built up according to the Data Driven Simulator approach. In this paper an applicative case will be analysed regarding the building of a simulation model of the Clinical Pathology Laboratory of the "G. Gaslini" Paediatric Institute, and .

***Keywords*—** ANOVA, Clinical Laboratory, Data Driven Simulation, Lagrange philosophy, Montecarlo.

I. INTRODUCTION

During the last years the interest in IT tool exploitation has grown up in the health and hospital field to carry out an effective and efficient management which must then combine the service quality given to the customer to the better and better use of the available economic funds, by limiting unnecessary losses. Among these tools the simulation has a primary importance role, since it allows achieving a virtual model of the examined reality and testing on it the effect of eventual changes on the system, such as for example lay out changes, purchase of new tools, etc.[1]

The simulation systems used for the hospital management usually show the same concepts which we can found in the industrial system simulation [2]. Moreover the simulation is an extremely useful tool to model the uncertainty, which is the main characteristic of the disease making it very suitable to the health system modelling [3]. The main problem that we meet in the health field compared, for example, to the industrial one is the high presence of the human intervention, which is lead by subject logics [4]. That is why in our case we decided to make use of a logic model in spite of a physical classical model.

Manuscript received April 9, 2007: Revised version received Nov. 24, 2007.

R. Revetria is with the DIPTEM, University of Genoa, via all'Opera Pia 15, 16145 Genoa, Italy (corresponding author to provide phone: 0039 320 7982156; fax:; e-mail: roberto.revetria@unige.it).

P. Giribone is with the DIPTEM, University of Genoa, via all'Opera Pia 15, 16145 Genoa, Italy (e-mail: piero@itim.unige.it).

F. Oliva is with the DIPTEM, University of Genoa, via all'Opera Pia 15, 16145 Genoa, Italy (e-mail: francesca.oliva@unige.it).

M.Schenone is with the DISPEA, Politecnico di Torino, corso Duce degli Abruzzi 24, 10124 Torino (e-mail: Maurizio.schenone@polito.it).

II. THE ANALYSIS LABORATORY OF THE G.GASLINI PAEDIATRIC INSTITUTE

The target of our study is the carrying out of a simulation model for the Giannina Gaslini Paediatric Institute Analysis Central Laboratory of Genoa. The achievement of a simulation model for the analysis laboratory is fundamental to offer a better service to the customer, that is to the patient, since it can make possible to detect eventual "bottle necks" in the system and reduce the answer times of a field having a strategic importance in the diagnosis rapidity.

Our work includes in the field of a project providing a model achievement representing some laboratory sectors, particularly the Haematology, the Coagulation and part of the Clinical Chemical, also called "Integra", besides to the acceptance or pre-analytic phase. The total number of the existing sectors is well wider, but the mentioned ones can be assumed as representative of the Laboratory, because of their importance and the number of treated test tubes. The test tubes are the object that, sent by the wards or the ambulatory, is sorted at the reception and analysed in the laboratory sections.

The modelling of this kind of systems is particularly complex because of the high stochastic character and the great exploitation of human resources, those actions do not follow always a single and precise logic, and therefore they are hardly arranged in a scheme. Moreover, these logics would be in any case different from person to person. A further complicating element derives from the fact that there are alternative ways, times and different behaviours according to the test tube type, the test to carry out, the time bands and the analysis emergency level. In order to try to reproduce the system in the most realistic way possible, we have collected much information, through not only the observations in loco which have been organized according to a structured data collection, but also the interview to the expert personnel [5].

The target was to detect in the test tubes path the most significant instants, during which an event takes place involving the test tubes, and so dividing the total analysis time in different consecutive intervals, which we have called Delta. In this way for example it should be possible to detect the process sections in which the test tubes take more time.

The analysis in the 3 sectors which we took into consideration, that is Haematology, Coagulation and Clinical

Chemical, are carried out on different test tubes. On each test tube it is executed a single analysis typology. The request associated to a patient it is considered as concluded when all types of necessary exams will be executed and validated by a doctor. While it is simple to sum up the technical validation process path, which is linked to the machine operation, it is harder to detect the logics leading the doctor validation, since they are human and subject logics, changing form person to person. The haemachromes and Coagulation analysis follow the same path and they are analysed by the same doctor; on the contrary the Chemical analysis follows a different path and it is followed by its own doctor. The logics leading the doctor are not fixed and change from doctor to doctor. For this reason during our data collection phase in the Lab we were concentrated on the process section arriving to the result production and the technical validation, seen that we already knew that the following section would have been more hardly detectable with precision standard and that in any case it had already been studied on detail in the past using data coming from the Institute IT system. The data, in fact, had put into evidence a very high variability, and therefore it had been hard to detect an exact distribution according to the data collected during the sampling. The doctor validation part can be therefore introduced later in the model using this big quantity of already analysed and validated data, coming from the lab database.

The pre-analytic phase consists in the sorting of the test tubes and the other material coming from the different wards or the ambulatory. The samples coming from the wards arrive to the lab inside plastic boxes which are ranged on a stand out of the sorting room. On the stand there is a paper on which the person charged with the transportation marks the arrival time, so as to have a FIFO type management. For some samples it is carried out the manual check by the operator. In particular they are samples which will be sent to other sectors of the lab and not in those examined by us.

For the greater part of the samples, on the contrary, it is carried out the automatic check using a particular machine, the RSD 800, called “Sorter”, a machine which can be configured to automatically sort the test tubes and unplug those requiring it, able to process different kind of samples in an ordinate way. The machine RSD 800 is so called since it can have a sort capacity of 800 samples by hour.

III. The Lab Model

The model creation phase makes reference to the conceptual model creation, and concerns the comprehension of the real system behaviour and the detection of the most important steps to reproduce in the simulator. The building of a flow chart showing as the system operates (figure 1), eases a lot the comprehension of the concerned variables and the way they can interact among them. In our case the necessary schemes displayed in the following pages have been drawn up with the Microsoft Office Visio software. Our schemes belong to the basic flow chart category [8][9]. In our field, for example, we must face a highly stochastic reality, complicated by the great human intervention and by the fact that the test tubes can be different among them, come from different ward, require different analysis, that is tests for different analysis, and can have a different emergency degree. Each test tube is destined to a single lab sector, but on it different analysis of the same typology can be carried out. Therefore we decided to maintain a low detail degree, by observing the path of the test tube from the beginning to the end and by taking notice of the time spent in the different sections, without detailing the operation of the different machine logics. An other important things to notice is that for us the machines used to carry out the different analysis are seen as black boxes, we take into consideration the test tube loading time and the final result production time, without worrying of what which takes place inside, without considering how many and what kind of test are carried out on each test tube. Our modelling philosophy is included then in the Lagrange philosophy category, in fact we follow the singular test tube along the whole path through the different machines. Particularly we have introduced along the path some “Check Points” that is control points where we took notice of our sample passage.

The first check point which we had considered is the arrival to the stand of the boxes containing the samples, while the last is the time where the analysis is concluded and the technician validates it.

The time between a check point and the other constitutes indeed a particular “Delta”, that is one of the time interval in which the path has been divided, and for which we have tried to find some distributions. Particularly, the more interesting Delta are two:

Delta0 → the time between the test tube arrival to the stand inside the boxes coming from the wards and the check in time.

Delta1 → the time between the check in time and the technical validation.

Wanting later to keep into consideration also the doctor validation process, it is necessary to introduce another check point, and the time between the technical validation and the medical one will be called Delta3. Delta2 will be on the contrary the time between the check and the medical validation.

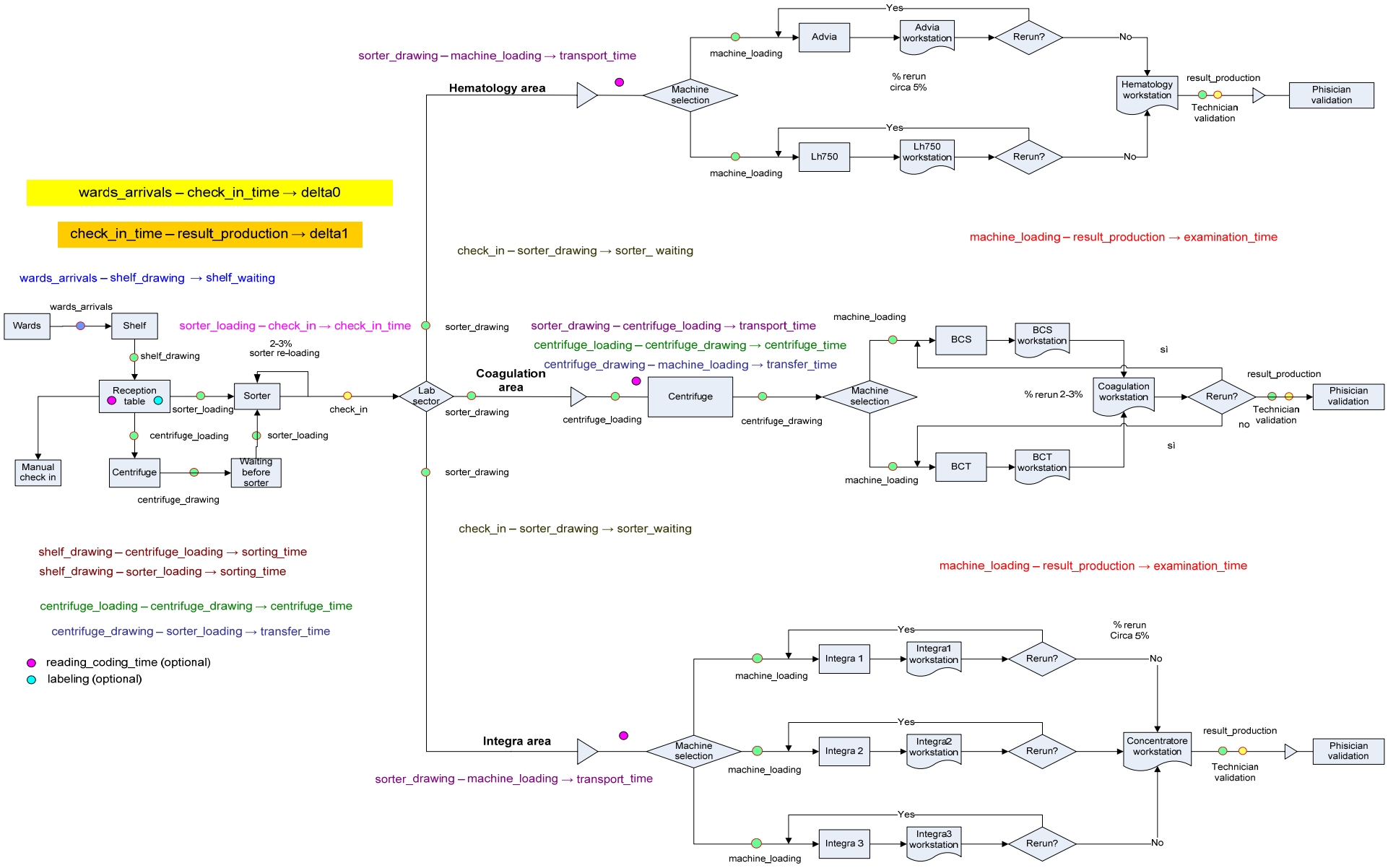

Figure 1: The flow-chart of the system

IV. DATA COLLECTION AND PROCESSING

This step is strictly related to the conceptual model creation phase, since the type and quantity of data to collect depend on the assumed model complexity and the main analysis targets. The data which are necessary to carry out the analysis have been collected on site by personally observing the process flow and by interviewing the lab technicians, that are people working in the system from many time and knowing almost all of it and who are therefore the only people able to giving precious information. The Gaslini Institute reached a high technological level and the sample flux is recorded in their archives; in fact for as concern the laboratory, some check points located in the most process critical areas are monitored by a data processing system, so information are collected on each test tube.

Num. Observation	Average Value	Probability
9	27,7	0,30
7	37,3	0,23
7	46,9	0,23
3	56,6	0,10
2	66,2	0,07
2	75,8	0,07

Table I – table example of distribution –Delat1 - Chemicals

Particularly we record the times relevant to the check in time, to the technical validation and the medical validation. We have not IT path of the pre-analytic phase before the check-in, then this information should have been collected personally. The target was to obtain some frequency distributions to introduce in the simulator: the data collection phase took place from February to April 2006.

Having to choose how many data collect we decided to have 30 of them for each Delta. 30 is in fact a significant number in the statistic field, since it is the established limit among the small and big samples, since it is considered a sufficient number for which we can be confident that the normal approximation is valid, according to the central limit theorem.

There has been some useful information but more difficult to take through the direct sampling (percentage of the rejected analysis, percentages of rerun), then the problem has been solved through interviews to the lab technical personnel.

As first approach parameters, on the times available for the different Delta, for the waiting time on the stand or for the inter-arrival data in the different days, we calculated the standard average and deviation.

Later, for each data assembly, we took the relative frequency histograms.

All the studied parameters and the graphics have been realized using the minutes as measure unit.

From the data collection, we took the separated data for each Delta and different kind of analysis (haemachromes, coagulation and chemical). Therefore it was logic to think that some data could be unified, but to be able to state this more precisely, we carried out suitable analysis using the ANOVA methods and a classification method (monovalent analysis). After having obtained the confirmation from those about the "bounty" of the aggregation, we unified the data [10][11].

Really, in this case the aggregation of some data was the better choice, since we did not dispose of quite samples to be able to carry out a precise analysis and diversified according to the different factors, such as for example the number of operators, the days or the time bands. In any case, to be sure of not doing an excessive approximation, we carried out monovalent analysis according to the factor that we considered more possibly influent, even if really these analyses, which have carried out on so few data and in a variable number among the different treatments, could not be very precise.

The data processing results are displayed in the tables (table I) in which the columns show in the order the lower and upper extremes of the different classes, the absolute and absolute cumulated frequencies, the interval average value and the relative and cumulated frequencies. From those we take a high variability, since the other classes, also the more extreme ones, show non negligible frequencies. Moreover, the possible value range is often wider. By observing the graph, we can notice that there is not a very predominant single class compared to the other, and this why, as we have many times explained, the process has a high variability, due greatly to the human intervention.

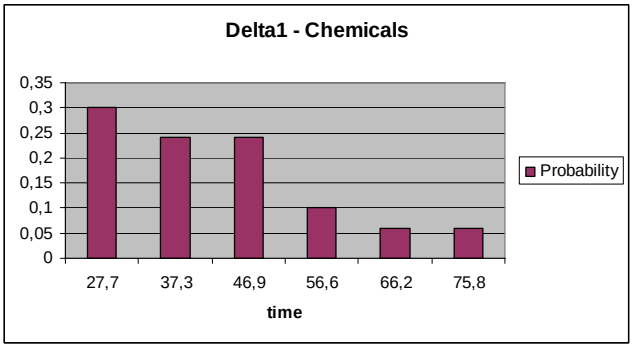

Figure 2 – example of histograms –chemicals

This other variability, in some cases greater and in other smaller, is in any case a characteristic of all the found distributions.

V. THE SIMULATOR AND THE DATABASE

After having concluded the data collection, the simulator has been built using the Arena Rockwell 8.0 software. It is an "abnormal" model compared to the standard ones that usually are built by organizing before the process scheme and then collecting the data to introduce in it. As we have already said, in the concerned case this method is not suitable, given the high system variability: we have therefore decided to follow a different approach, that is we have before collected the data relevant to the test tube process time, and then we realized a simulator such to reflect the collected data. It is not a physical model, but a logic one reproducing the test tube crossing time

recalling the different Delta times belonging to its path from a database connected with the simulator. Therefore we can say that the simulator is built up according to the collected data, the approach defined as "Data Driven Simulation"[12][13], through which the simulator, in its operation, interrogates a database containing all the necessary information: the big advantage is that the model can be easily reconfigured by simply changing the data.

To reproduce the detected frequency histograms, it follows the Montecarlo method logic by taking a random number between 0 and 1 representing the cumulated probability associated to one of the frequency histogram classes previously obtained "on site", and that it is then chosen during the simulation in progress [14][15]. Through this proceeding of time extraction, the different check points along the test tube path are generated [16].

The model realized in Arena is able to connect through Visual Basic for Application to a database built with Microsoft Access. From this database the model extracts the necessary input data and import also the simulation output data in the database. Therefore the database has been prepared with a series of tables containing the necessary information to realize the simulation and obviously the data relevant to the histograms to reproduce.

A. Simulator Check and Validation

After having realized the simulator, it has been carried out a check and validation phase of the same, to asses the correspondence of the model to the real system: the output generated by the simulation have been then compared with the historical data collected during the sampling phase [17][18].

The check test have been carried out through monovalent analysis on the total times Delta0 and Delta1, which are particularly critical since they have been obtained by the addiction of more intervals, each characterised by its own variability.

From a monovalent analysis point of view, we tested with a significance $\alpha = 0,05$ level, the hypothesis H0 (equivalence of the treatment averages) for:

Simulated data

Real data.

SS		D.o.F.	MS		F	
Total	13482,6	133			F_0	0,03
Tr.	3,3	1	Tr.	3,32	$F_{0.05,1,132}$	3,84
Err.	13479,3	132	Err.	102,12		

Table II: ANOVA Table "Delta0 – haemo&coa"

To have a balanced trial, we choose to work with a numerousness sample equal to that of the real available data, by obtaining for all the different wards the acceptability of the null hypothesis (table II).

The following step has been the Mean Square Pure Error analysis and the averages convergence, so as to verify if the first tends to zero and the second, after a suitable number of run, tends to stabilize near to the real data average.

For the all analysed intervals the MSPE tends to stabilize around the zero (figure 3), whilst for the averages we have remarked some problems, since they tend to stabilize on values different from the expected ones, we analysed the system to understand the causes of this abnormal condition.

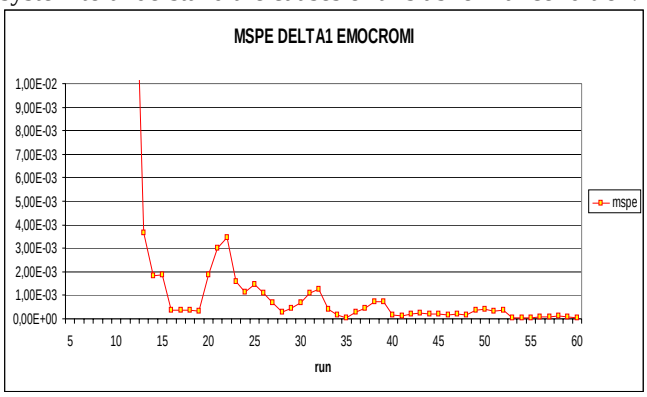

Figure 3: MSPE trend concerning the Delta1 in the haemachrome ward

The problems remarked are the following ones:

In some cases the sampling has been carried out on incomplete data paths, for which the "real" data are missing on total Delta which are obtained as addition of the singular characters, involving an increase of the total variability.

The data concerning the different test tube typology waiting time on the stand have been divided into 4 different distributions, each of them has a different average compared to that obtained only from the data relevant to the whole path, used for the simulator.

In the case of Delta1 the values of the simulator are increased by the presence of the estimated rerun percentages.

Then, really, the problems derive from the comparison of not homogeneous data. The Re-testing of the simulator after having eliminated the remarked differences, has put into evidence a substantial stabilisation of the average value compared to that calculated with the real data.

VI. Analysis of Time Intervals and Scenarios

The validated simulator has been used to analyse how mach time the test tubes spend in every step of their path inside the lab.

Every sector of the lab has been studied separately, and the different times can be observed divided step by step or grouped together in the interval Delta0, Delta1 e Delta3.

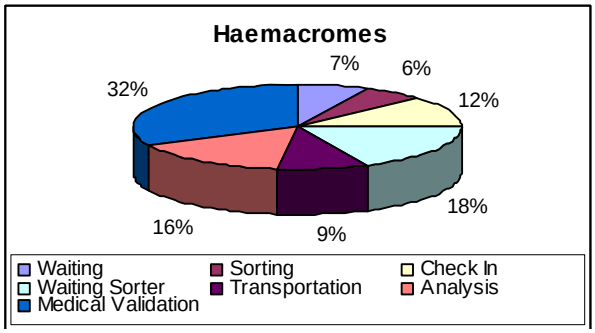

Figure 4: time analysis of the different step in the Haema sector

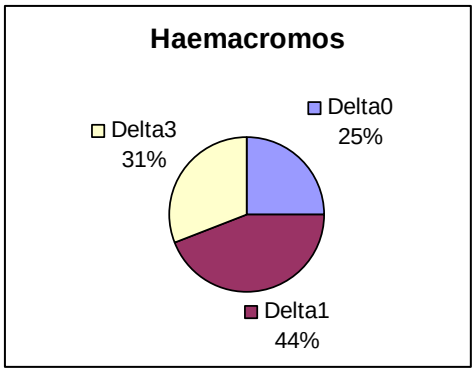

Figure 5: time analysis of the different intervals in the Haema sector.

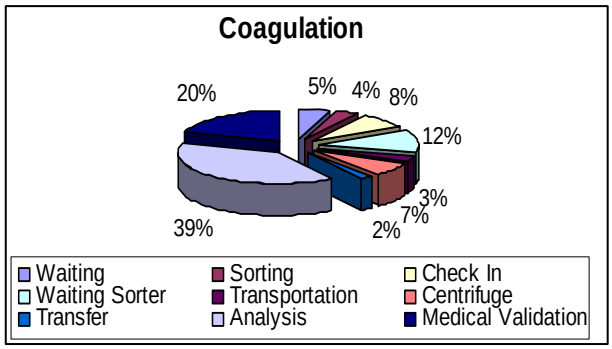

Figure 6: time analysis of the different step in the Coa sector

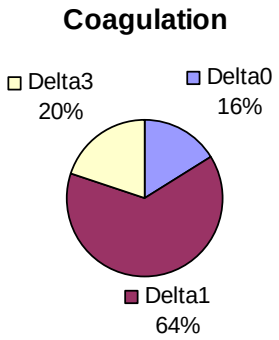

Figure 7: time analysis of the different intervals in the Coa sector.

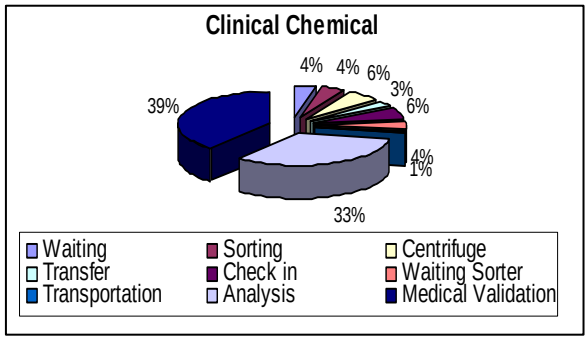

Figure 8: time analysis of the different step in the Chemical sector

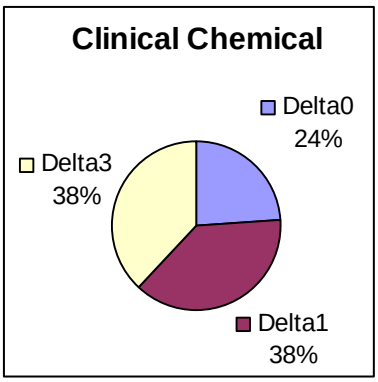

Figure 9: time analysis of the different intervals in the Chemical sector.

In all the sectors most of time is spent in the Analysis and in the Medical Validation. The first time in quite difficult to reduce, because it is connected to the working time of the different machines. The second one, that derives from the personal logic of every doctor, could probably be optimized, but this optimization is difficult because it depends completely on human behaviour.

If the overall time of each sector has to be reduced, the best solution is to work on the Waiting Time, that is not the longest, but that can be shorted avoiding high costs, just rationalizing the process. Some projects about the reduction of the Waiting and Transfer Times is on progress.

After this study, an analysis of different scenarios has been carried out to understand if the lab is able to manage the actual load and a possible future increase of work. The lab technicians have been interviewed to understand the actual load of the different sectors: according to them in an average day there are 45 boxes coming from the wards and each box has inside:

5 test tubes for the Haemacromes sector
2 test tubes for the Coagulation sector
9 test tube for the Clinical Chemical sector

These boxes arrive in the lab from 8.00 to 10.30 a.m., and after that the test tubes from the ambulatory come, that are on average:

40 test tubes for the Haemacromes sector
10 test tubes for the Coagulation sector
40 test tube for the Clinical Chemical sector

These data are used as input for the simulator: as output it produces the average number of test tube inside every machine in the different time of the day. Remember that we are working in a logical model, so the system is counting the number of test tubes that are or are travelled inside every black boxes during the observation time.

For example, the average load of the machine Advia can be observed in Figure 10:

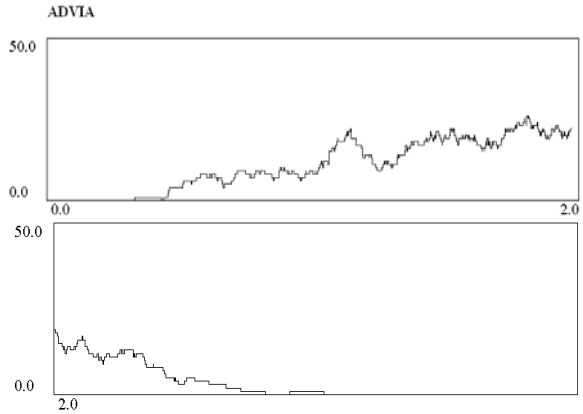

Figure 10: working load of the Advia from 8.00 to 12.00 a.m.

The maximum load of this machine (considered as a black box) is of 160 test tubes (10 in process and 150 waiting to be processed).

As it can be seen in figure 12, the number of test tubes inside the Advia is widely lower then its maximum load, and the other machines have more or less the same behaviour.

Then a new scenario have been tested doubling the load for every machines, but, as it intuitive, it does not create any problem or bottleneck, because all of them have an high availability: as example, the new working load for the Advia is shown in Figure 11

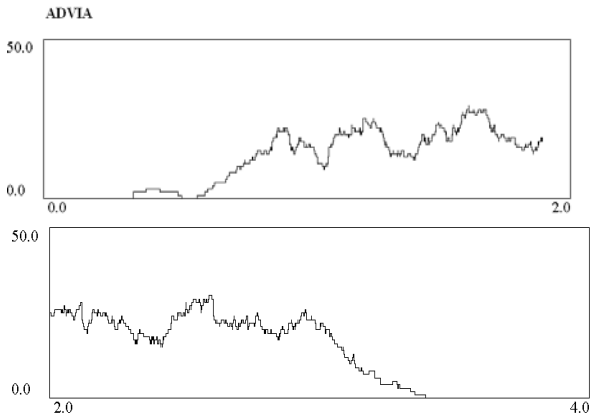

Figure 11: working load of the Advia from 8.00 to 12.00 a.m. in the new scenario

This means that the lab will be able to manage a possible future increasing load.

VII. Conclusions

The results obtained at the end of the model validation show as, in a so variable system, with a high level of human resource intervention, whose actions do not follow a constant and precise logic, a classic type simulator is not the most suitable choice. It is impossible to correctly describe and build the lab system through established waiting times, fixed loading logics or constant machine analysis times that in fact do not exist, so certain logics to implement in a simulator do not exist.

For this reason we choose to follow a different approach, following the test tube path and building frequency histograms for each of the time intervals the process has been divided in. In this way there are not rigid times for the different operations, but time extracted with a certain probability from a histogram, following a Monte Carlo method. Once the model has been validated, it has been used to find out the impact of every step on the overall time and to check if the lab will be able to manage an increasing load. The simulator is built according to the collected data, introduced in a suitable database: we talk about a Data Driven Simulation. It is not a classic physical model, for whose realisation established rules would be necessary, but a logic model. This kind of approach can be considered as generalizable and exportable also in other contexts. Particularly, it can be applied in all those cases in which it is impossible to define fixed and steady rules to implement in a model. This can be verified not only in the medical field but also in other fields, in all those cases in which we have a great human resource intervention.

Acknowledgment

The authors wish to thank the Prof. Dr. Giovanni Melioli and the Dr. Roberto Bologna for their extraordinary support an cooperation during the data collection and model definition phases.

References

[1] Connelly DP, Willard KE. Monte Carlo simulation and the clinical laboratory. Arch Pathol Lab Med. 1989 Jul;113(7):750-7.

[2] Brender J, Magdal U, Wiegell B, Schioler T, McNair P. Problem-oriented management of laboratory work through dynamic test scheduling. Clin Chim Acta. 1993 Dec 15;222(1-2):57-69.

[3] Josè A. Sepulveda, William J. Thompson, Felipe F. Baesler, Maria I. Alvarez, Lonnie E. Cahoon, The use of simulation for process improvement in a cancer treatment center. Proceeding of the 1999 Winter simulation conference

[4] Buffone GJ, Moreau DR .Laboratory computing--process and information management supporting high-quality, cost-effective healthcare. Clin Chem. 1995 Sep;41(9):1338-44

[5] Julie C. Lowery, Getting started in simulation in healthcare. Proceeding of the 1998 Winter Simulation Conference

[6] O.Balci, Principles and Techniques of Simulation Validation, Verification, and testing, Proceeding of the 1995 Winter Simulation Conference

[7] A.M.Law, W.David Kelton, Simulation Modeling & Analysis, Second Edition McGraw-Hill International Editions (1991)

[8] J.Banks, J.s.Carson, II B.L.Nelson, Discrete-event system simulation, Second Edition N.J.:Prentice-Hall (1996)

[9] P.A.Fishwick, Simulation Model Design and Execution, Upper Saddle River, N.J.:Prentice Hall (1995)

[10] Peck, Olsen, Devore, Introduction to Statistics and Data Analysis. Duxbury Press (2000)

[11] Lamberto Soliani, Fondamenti di statistica applicata, edizione settembre 2002)

[12] Bing Cao, Richard Farr, Mike Byrne, James Tannock, Data - driven simulation of the extended enterprise. 18th International Conference on Production Research.

[13] Solomatine, D.P., Data-driven modelling: paradigm, methods, experiences, Proc. 5th International Conference on Hydroinformatics (2002)

[14] Joy Woller, The Basics of Monte Carlo Simulations, University of Nebraska – Lincoln (1996)

[15] B.A. Wichmann, I.D.Hill AS 183: An efficient and portable pseudo-random number generator, Applied Statistics (1982)
[16] D. P. Landau and K. Binder, A guide to Monte Carlo simulations in statistical physics, Cambridge University Press (2000)
[17] Jack P.C. Kleijnen, Validation of models: Statistical Techniques and Data Availability, Tilburg University (1999)
[18] Robert G.Sargent, Verification and validation of simulation models, Proceeding of the 30th conference on Winter simulation Sheldon