Simulation Modelling Practice and Theory 104 (2020) 102149

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

journal homepage: www.elsevier.com/locate/simpat

Process mining as support to simulation modeling: A hospital-based case study

Oscar Tamburis[a,*], Christian Esposito[b]

[a] Department of Veterinary Medicine and Animal Productions, University of Naples Federico II, Via Delpino 1, 80137 Naples, Italy
[b] Department of Computer Science, University of Salerno, 84084 Fisciano, SA, Italy

ARTICLE INFO

Keywords:
Discrete event simulation
Healthcare
Petri-nets
Process Mining
ProM6
Simul8

ABSTRACT

The purpose of this paper is to show how the knowledge of Process Mining techniques can provide a robust premise to build a Discrete Event Simulation (DES) model of a healthcare process. In order to analyze some specific processes of an ophthalmology ward of a large Italian hospital, ProM6 framework was implemented, which supports process mining techniques in form of plug-ins; the plug-ins process data, in form of Event Log, in order to extract information about the process. The DES model based on such information was run via a commercial tool, Simul8, which allows building sophisticated process models. A timely algorithm was then deployed to adapt the ProM6 information to the DES model. Following, a conformance analysis was conducted by comparing the original Event Log, and the Simulation tool data, taking into account the main case-related model attributes (routing probabilities profile, time perspective, and resources perspective). The paper aims at developing the line of inquiry for what concerns the deployment of approaches to set forth the link between Process mining and Simulation modeling.

1. Introduction and background

Within the rising tide of Healthcare 4.0 [1-2] an increasing interest is currently witnessed for business processes simulation, because of the possibility to analyze the behavior of processes in any field and to know how to improve them. As the variability of a real process cannot be captured and represented via a deterministic behavior, it becomes therefore crucial to deploy a stochastic behavior [3-4]. More specifically, a Discrete Event Simulation (DES) model is a stochastic model through which a system, along its workflows, is modelled as a network of queues and activities. The activities are presented as a discrete sequence of events in time, interspersed with queues. The use of simulation techniques has rapidly grown in recent years in the healthcare sector, as support to the Business Process Management (BPM) [5] methodology especially for what concerns patient-flow management, as well as resources management and correlated activities [6]. DES modeling presents some clear advantages [see e.g. 7], such as:

- Time periods manipulation – Simulation provides the ability to speed up or slow down time flows for evaluation purposes. Many simulation packages are capable for instance to simulate a whole year's production within a couple of minutes, providing the user with a quick access to a large simulation period.
- Systems functioning understanding – Simulation is especially useful when systems cannot be viewed or observed in their entirety. Managers can use simulation to reconstruct a system by making clear which are its components, how they work, and how they are interrelated, in order to gain insights and learning about the system's interdependencies as well as interconnectedness features.

* Corresponding author.
E-mail addresses: oscar.tamburis@unina.it (O. Tamburis), esposito@unisa.it (C. Esposito).

https://doi.org/10.1016/j.simpat.2020.102149
Received 2 March 2020; Received in revised form 8 June 2020; Accepted 4 July 2020
Available online 08 July 2020

- Problem identification – Most modern systems are complex in nature. Simulation provides a way to explore them for the identification of occurring issues. Only by understanding all of the interdependent variables it becomes possible to identify the true source of the problem. This results in less time spent for solving the problem itself, than just trying to remedy its symptoms.
- Bottleneck analysis – The identification of bottlenecks by means of a simulation makes possible to quickly evaluate methods for addressing that specific issue.
- Preparation through 'what-if' analysis – A simulation analysis turns as helpful to figure out alternative work hypotheses in order to highlight possible modifications to the system's response under varying operating conditions, so as to improve its overall performances.
- Decision-making – The process analysis allows the user to evaluate, compare and optimize alternative designs and plans of simulation models. Simulation should be used when the consequences of a proposed action, plan or design cannot be directly and immediately observed (i.e., the consequences are delayed in time and/or dispersed in space) and/or it is simply impractical or prohibitively expensive to test the alternatives directly.

Deploying a DES-based system makes it possible to represent the real system from different points of view and to analyze the behavior and the characteristics of the system itself. The model so obtained is a representation of the reality built to provide answers to specific questions, in a time faster than real, and its study leads eventually to solutions whose reliability increases with the accuracy of the model itself. Nonetheless, issues exist concerning the treatment of the data, the building time of a simulation model, the knowledge of the attributes of the process analyzed, and the validation of the model itself [8-9]. Moreover, BPM life-cycle also requires to automatically interpret information stored in event logs and use this to discover, check, and enhance models capable of describing what is really going on in the business process. The use of process mining techniques (PM), as a process management technique that allows for the analysis of business processes based on event logs to address the mentioned issues, has been largely investigated in the literature [10-15]. PM allows gaining insight into various aspects, such as the process (or control flow) perspective, as well as the rate of resource involvement can be discovered (organizational perspective) and the time-related performance of a process can be investigated (time perspective) [16-18].

The present paper makes the first attempt of how DES model building can be coped with via the implementation of Process Mining techniques, in particular to gain insights into healthcare processes [19-21]. In particular, a DES Model related to the 'cataract process' of the ophthalmology ward of a large Italian Hospital was built, starting from the event log data stored in Process Mining tools, and the conformance between the two was tested by means of a timely algorithm. The DES model was built with Simul8, a commercial tool that allows the creation of sophisticated process models, while data were analyzed by means of ProM6 (an extensible, platform-independent framework as it is implemented in Java that supports a wide variety of process mining techniques in the form of plug-ins, which typically represent the implementation of a mining algorithm). The research originated from the need to find a trade-off between the clear advantages that a DES approach can provide to the process management domain (especially for what concerns the healthcare field, which still suffers from organizational and cultural barriers against a wide implementation of patient-centered technologies), and the need to extract those very data the simulation itself relies upon. The experimentation meant therefore to find out the connections between ProM6 and Simul8, so as to get to an effective analysis of the process via the simulation tool. On the other side, the mere extraction of event logs cannot work the mentioned issues out, although the necessary input can be provided towards a modern perspective of the medical professions. An experience in a hospital environment (that can be seen as a sort of controllable "black box") is therefore the necessary step before replicating the approach in the territory outside the hospital, where the presence of many different actors that use different healthcare systems/devices, is highly likely to cause a dearth of the overall coordination that makes possible to design a reliable care path for the patient [22]. The paper is organized as follows: after the introduction, in Section 2 some principles of Process Mining are introduced, and the two software the research seeks to find a connection between are described; the main steps of the control flow of the process are shown in Section 3, while in Section 4 the results of the conformance analysis between the original log and the DES model for the case study are described. Finally, some concluding remarks are reported.

2. Frameworks deployment

A business process comprises a set of activities performed in coordination within an organizational and technical environment and meant to jointly achieve a business goal. A business process model deals with a set of activity models and execution constraints between them. A business process instance represents instead a concrete case related to some company's operations and is composed in turn by activity instances [23]. An event log is the amount of information about a process, recorded sequentially starting from the case ID (the physical customer), up to the events, the related timestamps (i.e. the timings of when tasks were executed, or when they happened to be in a particular state) and other attributes as resources, costs, etc. Event logs store information as well about the originator of a task, i.e. who performed a specific task or initiated an event [24].

Process Mining (PM) aims at improving the extraction of knowledge from event logs recorded by an information system this by providing techniques and tools for discovering process, control, data, organizational, and social structures from them. In order to do that, it is essential to extract event logs from data sources (e.g., databases, transaction logs, audit trails, etc.). Three main classes of PM techniques are formalized in the literature [see e.g. 25,26], based on whether or not an original model of the process is available and, if so, how it is used:

- Discovery – No *a priori* model exists. Specific PM techniques can be used to extract knowledge from an event log to find a control-

flow model (in this case in the form of a BPMN model) that well describes the observed behavior.
- Conformance analysis – An a priori model exists, which is compared with the event log so that discrepancies between the log and the model are analyzed.
- Extension – An a priori model exists but the goal in this case, rather than checking conformance between the model and the original data, is to enrich the model with a new aspect or perspective not originally implemented.

Different process model languages can be used to capture the causal relationships between the steps, or activities, within the process. It is then possible to compare the discovered model with the event log by means of the deployment of specific parameters, among which: Fitness (indicates how much of the observed behavior is captured by – i.e., "fits" – the process model); Precision (a model is precise when it makes the user see more than what is visible in the log); Completeness (the model seeks to detect parallel tasks without the need to observe every possible interleaving between them); Generalization (the parameter addresses overly precise model. It refers to a model that adds cases eligible in the log); Simplicity (the simplest process model that can explain the behavior seen in the log, is the best model); Soundness (all process steps can be executed, and some satisfactory end state is always reachable) [27-28].

2.1. ProM environment

For the present research ProM6 was chosen as it makes available different PM algorithms, through which different process models can be obtained by the user. ProM6 supports different target formats as MXML (Mining extensible Markup Language) and XES (eXtensible Event Stream, the official IEEE-approved format for event log interchange). The latter was adopted in the research as format for both the original log, and the one extracted from the Simulation model to verify its accuracy.

The platform allows the iteration of a several number of plug-ins, which allow for the implementation of different functionalities, among which [29]: mining plug-ins (take an event log and produce a process model by implementing some mining algorithms that take into account different languages, like Petri Nets for Alpha algorithm, or heuristics for HM); analysis plug-ins (starting from event logs or process models, their performances span Petri Net analysis, Linear Temporal Logic (LTL) properties checking, decision point analysis, fitness analysis, etc.); export plug-ins (provide specific functionalities to export a model into a file in form of event-driven process chains (EPCs), Petri Nets (PN), spreadsheets, grouped XML log files, heuristics net, yet another workflow language (YAWL) files, etc.); import plug-ins (make it possible for ProM framework to work with a variety of existing systems like EPC Tools, ARIS, Protos, NetMiner, etc., as they can import a process model from a file and possibly use a log to identify the relevant objects in the model); conversion plug-ins (convert one type of model into another, such as YAWL models into EPC, or Petri Nets into BPMN models, etc.).

2.2. Simul8 context

Simul8 [30] is a commercial tool for Discrete Event Simulation that makes possible to build sophisticated simulation models. The software allows for the user to model a system and evaluating its operations, with specific focus on: (i) Productivity analysis – times and costs of production; (ii) Management of incoming/outgoing; (iii) Management of queues and waiting times; (iv) Resources optimization. The control flow of the process makes it possible to add some rules with the purpose to define which conditions determine the duration of the activities, the division of the workload between the resources, etc. Moreover, performance measures are automatically collected so that the user, other than having a thorough view of the entire process, is able as well to get precise numerical results from each part of it. Although Simul8 is based on a graphical approach to build organizational models, rather than relying on statistical data or programming, a bi-directional interface with Visual Basic is implemented as well, which leaves space for the creation of advanced features that might not be modelled by means of a simple graphical interface. The software is also characterized by its own simulation coding language called Visual Logic, which allows for the user to implement a detailed logic simulation.

3. Process model design

The research was aimed at obtaining the control flow of the cataract process, as treated at the cataract center of the academic clinic for ophthalmology within the healthcare organization involved. Cataract is treated by performing cataract surgery, and its desired clinical outcome is an improved vision. In a cataract surgery the original eye lens core is removed, but placing a new intra ocular lens (IOL) is nearly impossible, since there is no remaining structure to attach the new lens to [31]. IOLs have been improved in years, especially thanks to an innovative technique called phacoemulsification, through which the surgeon can remove cataracts with a much smaller incision [32]. The current standard procedure is to use phacoemulsification in combination with a foldable lens. The required pre-operative eye tests are scheduled on the same day of the consultation with a junior doctor, which is supervised by a medical specialist. The tests and consultation are both forms of pre-operative examination, while the surgery itself is an outpatient procedure. The day after the treatment there is either a consultation by phone or by means of a physical examination, in case the surgery was carried out namely by a medical specialist or by a junior doctor. After a few days there is a physical examination. The results from the first cataract treatment are taken into account when performing the second surgery, which is not necessary in every case. If the second eye is not treated, there is a final post-operative examination about one month after the first surgery. The data sources stored in the Hospital Information System (HIS) from which the event logs have been extracted describe the set of activities

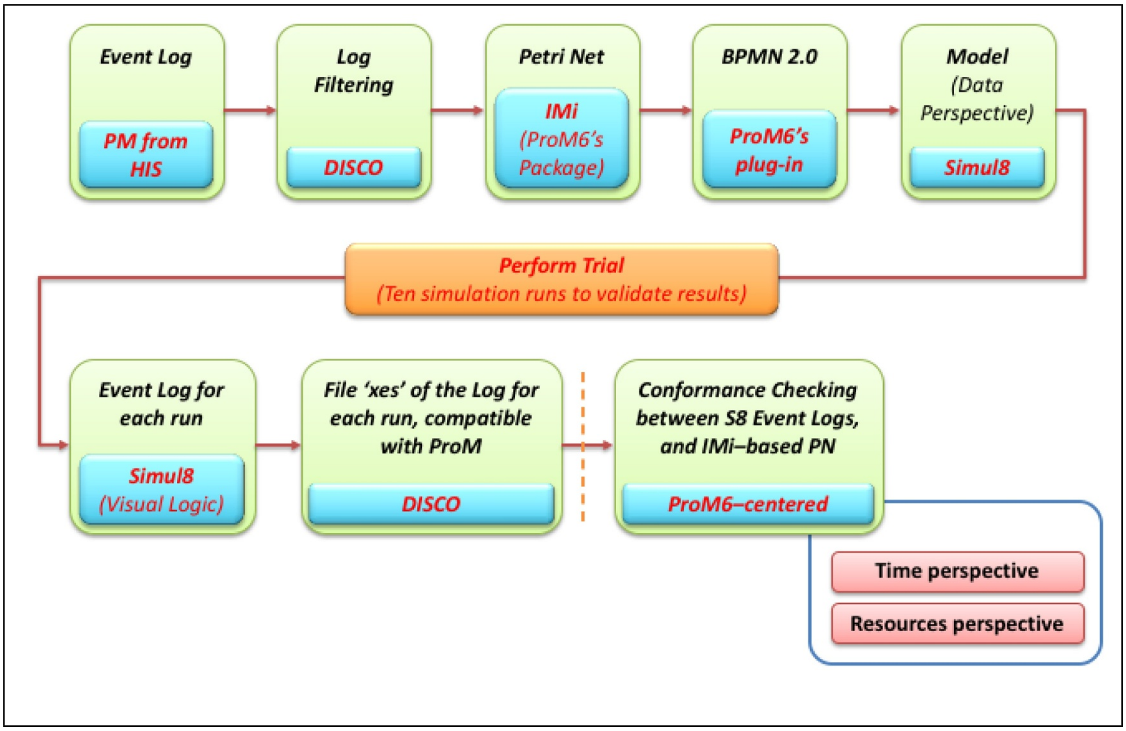

Fig. 1. Sequence of steps to build the control flow of the process.

performed in the ophthalmology area of the hospital, and relate to patients, events, timestamp, attributes which describe characteristics of patient's clinical admission and their ongoing (doctor specialist, date of admission, main and secondary diagnosis, treatment, etc.). Our research aimed at figuring out a process discovery algorithm, since it was necessary to construct a process model capable to represent a behavior observed in the event log. The main steps that characterize for our case the general algorithm of the project that links ProM6 and Simul8 environments are depicted in Fig. 1. For each steps the software implemented to perform the correspondent activity is indicated.

In order to build up the event log to analyze, only patients of the ophthalmology ward that had their process starting point after the March 29th at 11:30:00 CEST 2014 until the November 22nd at 10:30:00 CEST 2018, were taken into account [33]. The original Event log contains more than four thousand events and seventeen activities. Table 1 reports the main activities considered (both in Italian and English languages) as well as their position within the whole process organization.

In most real-life event logs, some paths are infrequently taken, or their traces only differ because of the occurrence of infrequent activities. Such logs contain infrequent behaviors that challenge the performances of discovery algorithms: if an infrequent behavior is included in the model, simplicity might be sacrificed; if it is excluded from the model, fitness instead might be sacrificed. Typically, the 80% of the observed behaviors can be explained via a model that is only 20%-sized than the model required to describe all possible behaviors [34]. To obtain such a model, a classical approach is to globally filter the log before discovering the model itself, i.e. to remove that data range that turns as not indispensable to understand the general process behavior. The Disco process mining tool was deployed to that end. In our case, the original event log was in a '.xes' format, so it could be directly imported in Disco, which gave back a preliminary process model. The analysis of the frequency of the process activities made possible to discover that most events occurred not so frequently in the process, or even in only one case: it was therefore decided to filter the log, by keeping only the performed services (called events) with absolute frequency greater than one hundred occurrences, as well as the patients (called

Table 1
Main 'Cataract Process' – Related Activities.

Activity (Italian)	Activity (English)	Refers to:
Topografia corneale	Eye scans	Preoperative examination
Biometria ottica	Optical biometry	Preoperative examination
Infermieristica di prima linea (pesante)	Heavy day time nursing	Services performed on the day of surgery
Estrazione lenti oculari. Impianto nuove lenti	Eye lens extraction and implanting a new lens	Services performed on the day of surgery
Tempo di taglio a sede singola	Making an incision in a surgical session in which only cataract is treated	Services performed on the day of surgery
Durata singola sessione chirurgica	Being in surgical session in which only cataract is treated	Services performed on the day of surgery

cases) whose sequence of activities had been shared by at least two cases. It was so possible to figure out the main sequence of the process activities. The filtered Event Log of the process featured the following statistics: 2789 events, 790 cases; 18 activities; 13.2 weeks as median value for case duration; 14.6 weeks as mean case duration; start activities on April 4th at 10:20:00 CEST 2014; end of activities on October 18th at 11:10:00 CEST 2018.

The filtered event log was then imported again in ProM. Since event logs related to clinical data are often incomplete, or affected by noise (e.g. due to missing data or incorrect timestamps), it was necessary to deploy an algorithm capable of returning a sound model in any case (that is free of deadlocks and other anomalies), well-handling infrequent behaviors, and providing a quick performance. The technique chosen that showed to be able to cope with infrequent behavior and large event logs while ensuring soundness, was an extension of the IM (inductive miner) approach, called Inductive Miner–infrequent (IMi), which aims at quickly discovering an 80% sound process model. IMi is implemented in the InductiveMiner package of the ProM framework [35]. Frequencies of traces and events, ignored by IM, are taken into account by IMi in order to distinguish frequent from infrequent behaviors. Compared to IM, models discovered by IMi feature a lower fitness, higher precision, equal generalization and comparable simplicity. It is also possible to compare IMi to several existing techniques according to the mentioned performance parameters. In any case, other approaches designed to ignore infrequent behaviors can return as well an 80% of the model, nonetheless they deliver lower performances as to other quality criteria: it is e.g. the case of genetic approaches [27] that present long run times, or heuristic approaches [36] that produce unsound models. IMi, as many of the abovementioned algorithms, can express the mined process model originated from the filtered log, in terms of a Petri Net (PN) [37], where the token is the patient which undergoes the cataract process [38]. The choice of PN depends on the one hand on the need to respect particular formal criteria such as fitness and generalization, and on the other hand to deploy a robust enough algorithm to treat noisy data. To that end, the parameter of noise threshold was set at 0.10 (the value 0.0 represents the perfect log fitness), as it was decided to represent only the main behavior of the process [39].

Since PN is not an executable language it was necessary to turn it into an executable one: in particular, it was chosen to use the Business Process Model and Notation (BPMN) standard [40], as Simul8 does not import directly Petri Nets, while it accepts, among the others, a BPMN 2.0 format. A conversion was then run [41] via the ProM's `Convert a petri-net into BPMN diagram' plug-in undergoing specific qualitative checks, i.e.: numbers and names of events (some activities could be lost in the importing, or the names could be changed); layout of the diagram (the position of the events must be the same). Once designed the BPMN diagram in Simul8, it was possible to understand the traces of the work items, corresponding to the abovementioned cases. Work items are called to respect the traces of the process with a certain probability; this is true if the structure is well done. In order to see the traces of the work items, once run the Simulation it was necessary to extract the event log of the Simul8 structure; this was made via the creation of a general code in Visual Logic (VL), which extracts for each work item the activity it has been processed in, along with the corresponding timestamp. The code structure must be general, so to make it possible to reuse it in every structure in Simul8. The code sections refer to the areas of the activity in which the VL code works; in particular, the first part (called Action Logic Area) deals with the timing of the work item entered in the activity; the second part (called Exit Logic Area) reports instead the exit time. The data about each work item were set in a spreadsheet, which actually represents the event log extracted from the Simul8 model. Columns represent (i): the lb_id, which is the label number associated with each work item; (ii) the specific activity; (iii) the timestamp; (iv) the indication of start and conclusion for each activity, with reference to the corresponding logic area. Each row represents one event of a

```
VL SECTION: Activity Action Logic
  SET row  =  row+1
  SET spreadsheet[1,row]  =  lb_id
  Get Time Work Item Entered Object    data
  Time Convert Simulation to Local    data ,  data1
  Format Local Time    data1  ,  data2
  SET spreadsheet[2,row]  =  Activity.Name
  SET spreadsheet[3,row]  =  data2
  SET spreadsheet[4,row]  =  "start"
VL SECTION: Activity On Exit Logic
  SET row  =  row+1
  SET spreadsheet[1,row]  =  lb_id
  Time Convert Simulation to Local    Simulation Time ,  data
  Format Local Time    data ,  data2
  SET spreadsheet[2,row]  =  Activity.Name
  SET spreadsheet[3,row]  =  data2
  SET spreadsheet[4,row]  =  "completed"
```

1	dagverpleging zwaar	10/03/2014 09:12:14	start
1	dagverpleging zwaar	10/03/2014 09:22:14	completed
1	Snijtijd enkelvoudige zitting	10/03/2014 09:22:14	start
1	Snijtijd enkelvoudige zitting	10/03/2014 09:32:14	completed
2	oog~IOL~master	10/03/2014 09:28:03	start
2	oog~IOL~master	10/03/2014 09:38:03	completed
1	OK Anker verrichtingen	10/03/2014 09:42:14	completed
3	eerste pkl, bezoek~administratief tarief	10/03/2014 09:40:05	start
3	eerste pkl, bezoek~administratief tarief	10/03/2014 09:50:05	completed
3	afspraak	10/03/2014 09:40:05	start
3	afspraak	10/03/2014 09:50:05	completed
4	oogl,~extracaps,extract, impl,lens a,o,k	10/03/2014 09:40:21	start

Fig. 2. Example of Visual Logic code (up) and of event log extracted from Simul8 (down).

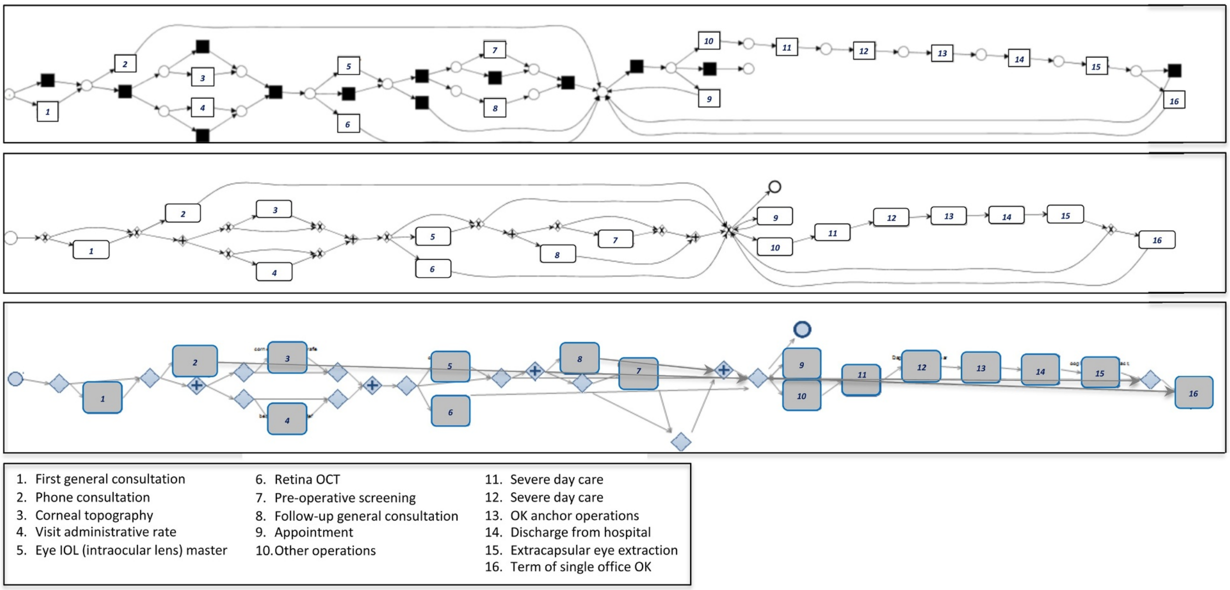

Fig. 3. Comparison between Inductive–Miner infrequent Petri-net of the cataract process (top/left); BPMN diagram obtained in ProM6 (middle); BPMN diagram imported in Simul8 (low/right).

case; by grouping all the rows that match a specific label of the work item, it is possible to extract a case. Fig. 2 shows both the VL code and a section of the related event log in form of spreadsheet.

In Fig. 3 the original PN, the BPMN conversion in ProM, and the BPMN imported in Simul8 environment are confronted, showing that both models respond to the requested checks. Nevertheless, the adaptation of the model in the Simulation tool is not trivial, and a validation process is needed, as described in the next sections.

3.1. Data perspective

In order to perform an as effective as possible conformance checking, Simul8 needed to be enriched with the data of the process. As a DES model is actually a stochastic model, in order to build a correct model it is critical to achieve for the process activities: (i) the clear definition of inputs and outputs of the actual system; (ii) the correct definition of the probability distribution related to the behavior of the system. Using ProM6 makes it possible to get good information of this kind: the ProM6 Log summary reports in fact all the information related to the system, such as: (i) absolute occurrences of the activities in the process; (ii) starting occurrences of the start of the activities in the process; (iii) percentage of the utilization of the resources in the process. In Simul8, the Routing in and Routing out features manage the coming-in/out of the token. In particular, for what concerns the routing out it is possible to define a route type, such as circle, percentage, label-dependent, uniform, or to use a visual logic interface. That being said, Simul8 allows the user to create in a specific routing out window the routing probability profile of each activity: this indicates the probability for the single activity to be performed (before or after other activities, or with a predefined duration) within a specific distribution of activities figured out to simulate the overall process under exam. In order to create the right distribution profile for each activity to be exported in the DES model, the BPMN model is therefore defined by parallel and exclusive gateways to represent different constructs of the process; for each possible construct, a formula was figured out to obtain the right routing probabilities profile of each activity.

On such bases, the routing probability is expressed in percentage by means of the P_{ai} value, where: ‘a’ represents the activity the user is taking the actual value from, and; ‘i’ represents one of the three cases that the user can find in the BPMN structure, namely: (i) transition gateway-parallel; (ii) transition gateway-activity (w/o back); (iii) transition gateway-activity (with back). Once obtained the P_{ai} for all the activities, it was possible to use them in the Simul8 model, so as to define the routing out percentages for each single activity. Fig. 4 shows some instances of the three possible cases the user can find in a BPMN model. The following formulae report examples for all the three cases, namely:

- Case 1 (parallel activities):

$$P_{a1} = start\ activity1\ \% + ...+start\ activity\ n\% = TOT\ parallel\ \% \tag{1}$$

- Case 2 (start activity). The start activity percentage can be found in the ProM log summary in the Start Events section; if no start percentage for that activity exists, it is possible to use the value contained in the All Events section:

$$P_{a2} = start\ activity\ \% \tag{2}$$

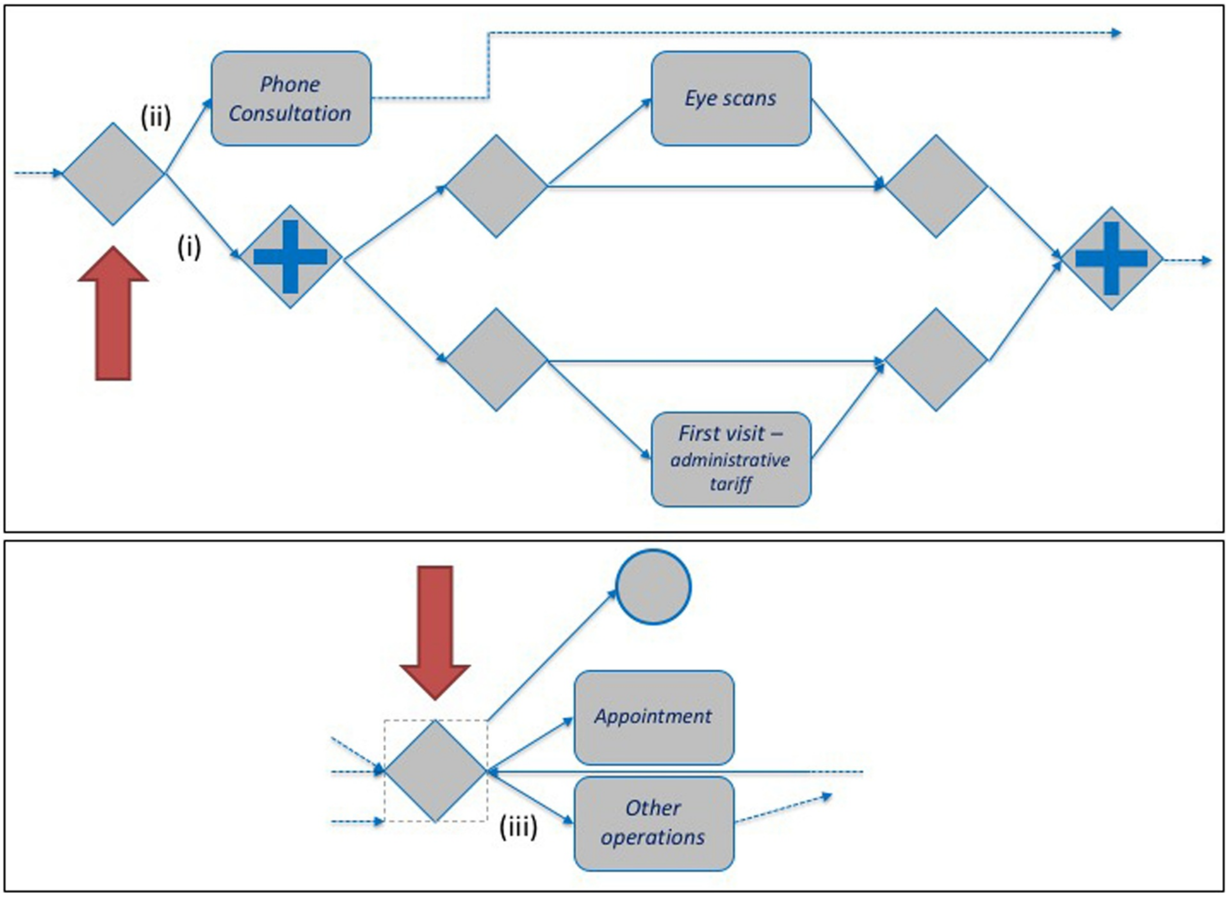

Fig. 4. The three possible cases the user can find in a BPMN model.

- Case 3 (number of occurrences): the following formula, used for instance to evaluate P_{end3}, $P_{interview3}$ and $P_{AN\ Anchor\ Operations3}$ (Fig. 3), features: 'n' = number of the absolute occurrences of the *i*-th process activity; 'ns' = number of the occurrences of start of the *i*-th process activity; the denominator of (3) is also called normalization parameter, as it is the sum of the Relative Occurrences for each *i*-th activity (ROi).

$$P_{a3} = \frac{n}{\sum_{n=1}^{\#\ of\ activities} (n - ns)} \tag{3}$$

There are two main ways to use these values in Simu8: the first is to export the value in an Excel file and import the data contained in this file in Simul8; the second is to put them manually in the gateways of the BPMN model, before importing it in Simul8 . For sake of simplicity we used the second way.

4. Conformance checking

Conformance checking is a PM technique that compares an existing process model with an event log of the same process [27]. In other words, it can be used to check if reality, as recorded in the log, conforms to the model and vice versa. Unlike process discovery, conformance checking takes both a model and event log as starting points. While conducting a conformance check the behavior of a process model and the behavior recorded in an event log are compared to find commonalities and discrepancies. Such analysis may result in global conformance measures (e.g. 85% of the cases in the event log can be replayed by the model) and local diagnostics (e.g. activity x was executed 15 times although this was not allowed according to the model). The interpretation of non-conformance depends on the purpose of the model. If the model is intended to be descriptive, then discrepancies between model and log point out that the model needs to be improved to capture reality in a better way. Some of the discrepancies found may highlight undesirable deviations from the original model: in this case, conformance checking signals the need for a better control of the process. Other discrepancies may reveal desirable deviations, e.g. workers may deviate from some predefined protocols to serve the customers better or to handle circumstances not foreseen by the process model. Most conformance checking techniques are based on the principle of Replay, which means that the event log is replayed on the process model. For example, while replaying an event log on a Petri net, one can count the number of missing and remaining tokens. Alternatively, one can try to optimize the mapping of traces onto models by introducing costs associated so as to skipping or swapping events in the log and/or model.

For what concerns Simul8, in order to obtain statistically significant results from the validation process the user is given the possibility to make a Trial, that is a series of runs of the simulation model resulted from the steps previously described. In our case, it was chosen to conduct a ten-simulations Trial to validate the results for the case study [42]. As it was possible for each run to create the related event log, a timely Visual Logic code was created similar to the one used to validate the structure and shown in Fig. 4. The only difference pertained to the need to introduce a variable "i" in the spreadsheet, which incremented the columns when shifting from one run of the simulation to another. Simul8 allows the user to eventually export this kind of spreadsheet directly in Excel. As a consequence, once obtained the spreadsheet, the .xls file of each run, as allowed format, was imported in the Disco tool: such step was necessary to obtain a '.xes' format, so as to get the ProM log summary of each event log of Simul8, containing both absolute and relative occurrences for each process activity.

In relation to what stated, as ProM6 does not feature any plug-ins capable of comparing directly two event logs, two different situations had to be coped with:

- Routing probabilities profile: a quantitative approach was implemented which featured the comparison between the original log and the Simul8 event logs produced in the Trial by means of a specific parameter;
- Time/Resource data (case-related model attributes): the 'Replay a Log on Petri-net for Conformance analysis' plug-in to replay the event log on the Petri Net model was implemented. In other words, the compatibility between the IMi Petri-Net model (obtained in ProM6 from the original data) and the event log (obtained from each Simulation run) was checked, in order to optimize the mapping of the log traces and, as a consequence, improve the alignment between the original log and the final DES model. The plug-in uses the Fitness as conformance parameter. According to the literature, a reliable alignment occurs for a Fitness parameter value comprised between 0.95 and 1 (the problem of a lower Fitness could only appear for hand-generated Petri Nets) [43].

4.1. Routing probabilities profile

As no specific plug-ins were in this case available to make a direct comparison, the quantitative approach implemented [44] was articulated as follows: (i) creation of the '.xes' format for each of the ten Simul8 run; (ii) import each run in ProM6, so as to obtain the mentioned ProM log summary; (iii) export the summary in HTML, so as to have the HTML file for each run-related Event Log; (iv) import the HTML files in Excel, so as to have the possibility as well to calculate, for instance, the average between the absolute occurrences of the same activities in the Trial; (v) composing a whole Simulation event log, which represents the average of the runs in the Trial, to be further compared with the original event log; (vi) compare this Log with the original one. In case of errors, conduct a comparison between each activity distributions profile to find out the gaps.

The comparison occurred then between the working time percentage of each activity in the original log and the average working time percentage of each activity in the Simul8 event log. The two-sample Kolmogorov-Smirnov (KS) test [45], which compares two data sets to decide whether they were sampled from population distributions with the same shape, was deployed via Matlab, considering a I type error $\alpha = 0{,}05$. Since the test results showed $p \approx 0{,}02 < \alpha$, it was possible to refuse the H_0 hypothesis (the result origins from random causes) and to affirm that the two distributions were from the same population. In this case, the mean absolute percentage error (MAPE) between the numbers corresponding to each activity was also calculated: an error, although not important (for example at the second decimal places), could in fact always affect a hand-driven importation of data in Simul8. MAPE is a measure of accuracy for constructing fitted time series values in statistics, specifically in trend estimation [46-47]. It is usually expressed as accuracy in percentage, defined by the following formula,

$$MAPE = \frac{1}{n} \cdot \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \tag{4}$$

where: 'At' = the actual value of the t-th activity; 'Ft' = the forecast value; 'n' = the total number of activities of the process. To obtain the percentage error, M is then multiplied for 100. If MAPE > 5% for a specific activity, this means that there is an error in the percentage transcription at the second decimal place for that activity.

4.2. Time perspective

Timing is a process attribute that affects all the components of the process. It is not possible to record an event log of a process without its timing view. Timestamps and frequencies can be used to learn probability distributions that adequately describe working times of the healthcare operators, in terms of waiting and service times. PM techniques show that the replay techniques used for conformance checking can be modified to add the time perspective to process models [39]. In our case, the 'Replay a Log on Petri-net for Conformance analysis' plug-in was implemented to replay the event log on the Petri-net model, so as to add the time perspective to process models. The plug-in gives back timing view tables as output, where time is expressed in average time, minimum and maximum value in the transaction between activities. The tables can be exported in Excel and used for a triangular distribution in the DES model: the upper, lower and modal correspond in fact to maximum, minimum and average values namely in the ProM6 table.

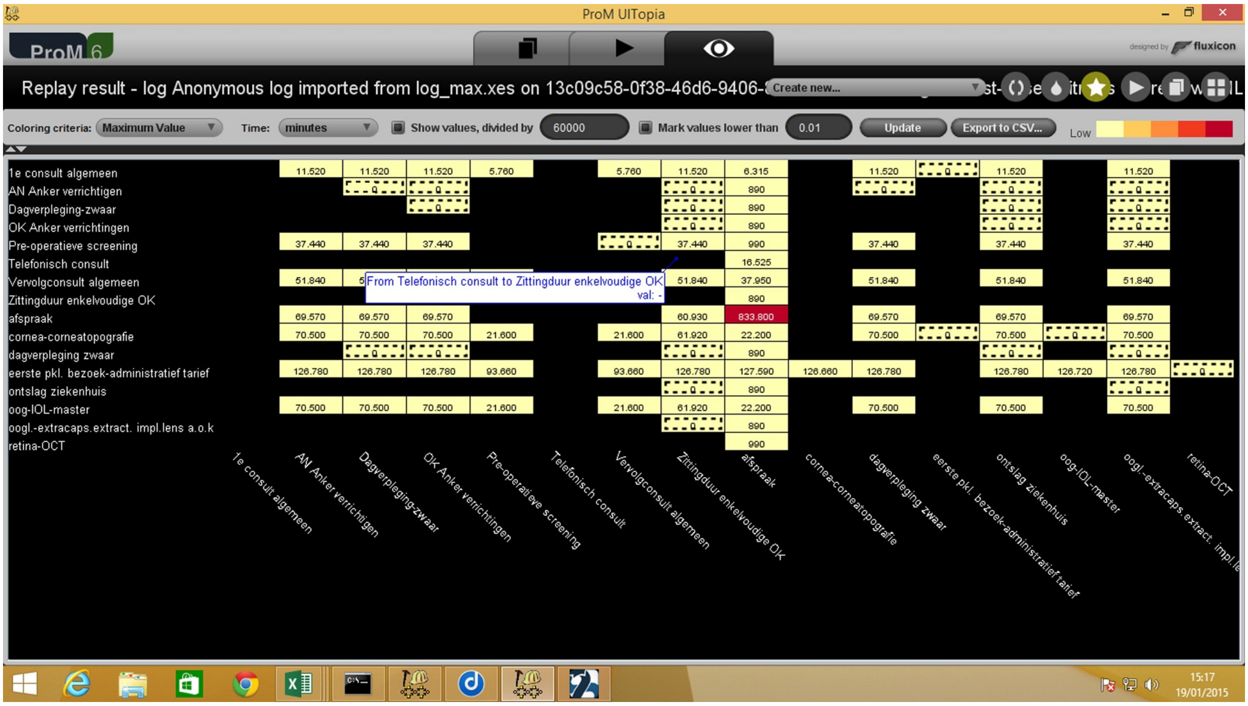

Fig. 5. Table of the maximum value of the time in the transitions.

Such distribution is needed to represent in Simul8 the transaction time between two activities. The Simul8 model had also to be enriched with the real start date (timestamps) of the process activities; this means that, at the end of the already described steps, the ‘Replay a Log on Petri-net for Conformance analysis’ plug-in, after being deployed to confront the IMi Petri-Net model and the Simulation-related event log, was also used to obtain the timing perspective tables for each Trial run, exportable as CSV file as allowed input of the ‘.xes’ format. It was so possible to calculate in Excel an average of the timings of the single run-related logs, which were compared with the original tables. In this case, the comparison was only qualitative, since the event log obtained from the original data of the cataract process only featured the Start Events of the process activities – thus missing End Events as well as activity durations.

If it is difficult to get to an exact knowledge of the activities’ durations in the process, it becomes accordingly difficult to use the ProM tables values to enrich the DES model. This is noteworthy, since usually the timing tables only feature the timestamp of the starting moment of the Events, along with the transaction time, i.e. the duration of the first activity, plus the time for the necessary transition to reach the following activity. This latter time value is not defined in the DES model, though it is important in the simulation to know the activity duration split from the transition duration. For instance, if the table contains the transaction time value between the activities ‘1st general consult’ and ‘first visit administrative rate’, this value represents the sum of the duration value of ‘1st general consult’, and the transition between the two activities; the timing value of this activity has therefore to be subtracted, in order to have the effective duration of the transition (see Fig. 5).

In any case, such approach only works if the event log is complete or if it is possible to know somehow the duration of the activities in the process. Of course, it is always possible to obtain data as to the single activity duration through the study of the literature or doing interviews directly to the resources (workers) in the organization, in order to extract the exact information needed.

4.3. Resources perspective

In a DES model it is possible to define the actors (called “resources”) who actually perform the activities, along with their characteristics, such as the overall workload, or the percentage of working time for a specific time period. Getting exact information concerning the resources is important, since the better the resource characterization is, the more realistic the DES model becomes. For such reason, it was important to investigate the workload-dependent behavior: traditionally, in fact, the so-called “Yerkes-Dodson Law of Arousal” suggests that a worker under time pressure may become more efficient and thus accomplish tasks in a faster way. However, if the pressure becomes too high, the worker's performance may degrade [48].

In our case the ProM's ‘Perform predictions of Business process features’ plug-in was also exploited, which made possible to discover some important rules lying underneath the process, i.e. ‘in which conditions a resource performs an activity?’, ‘who does what?’, or ‘how often?’. In this way ProM6 also provided information like the Correctly Classified Instances, Incorrectly Classified Instances, mean absolute square error, the total number of Instances and the Confusion matrix. The latter in particular contains the information regarding the number of resources that perform a specific activity. Rows represent the single resources, while columns

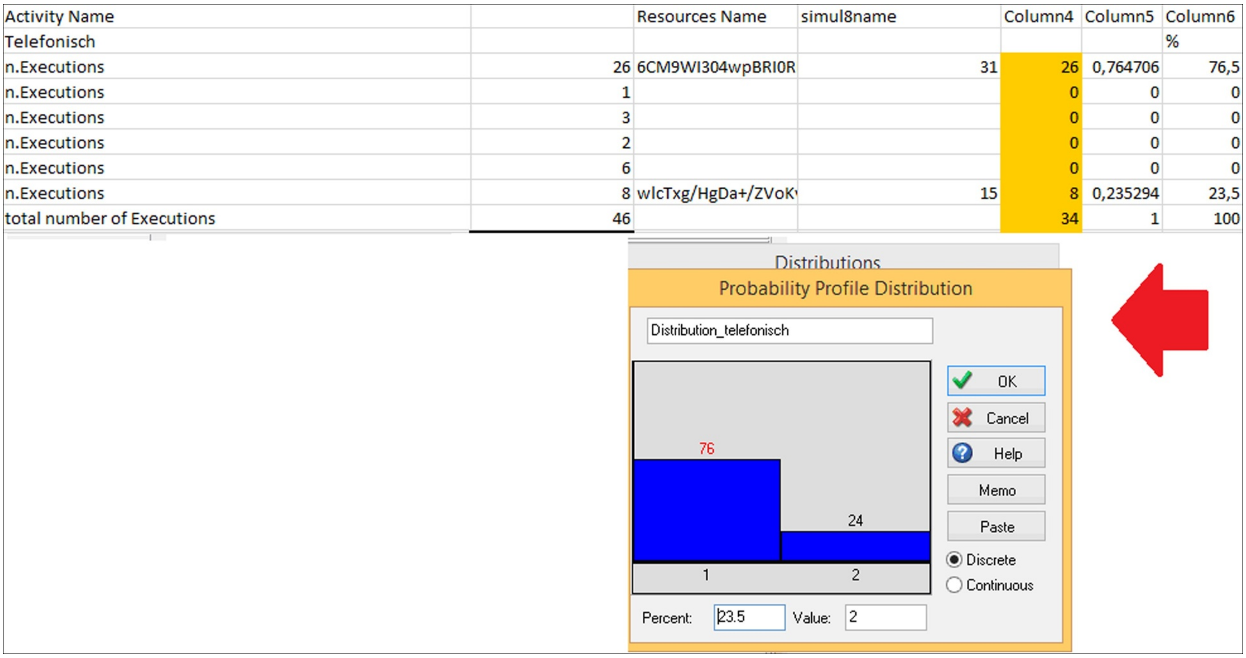

Fig. 6. Probability profile in Simul8 environment for the “Phone consultation” activity.

report the number of executions, as a function of the selected activity. For each row it is possible to calculate: ‘Xn executions’ = sum of the executions for each resource; ‘Xn executions *Av*’ = average number of executions of one resource for a specific activity. In formulae:

$$X_n\ executions\ Av = \sum_{resources} X\ executions\ (resource)/\#\ resources \tag{5}$$

The resource probability profile P_{ai}, in form of conditioned probability, is then calculated as follows:

$$P(resource|Activity) = \sum_{resources} X\ executions\ (resource)/ \sum_{resources} X\ executions\ (resources\ chosen) \tag{6}$$

A timely criterion was then adopted to determine the resources with the greater amount of work, in order to choose them before the evaluation of the P_{ai} in (6): in particular, the resources were chosen for the present research which featured a number of executions greater than the average of the total executions of the resources related to a specific activity. Of course, the choice of the criteria generally depends on the purposes of the analysis performed, see e.g. [12]. The information about the working percentages in function of the executions of the single resource as well as the total number of executions for that activity, were then treated in Excel, in order to obtain for each activity a ‘resource perspective summary’. Following, in Simul8 environment a distribution profile of the work performed by the resources for a single activity was defined, and a series of labels were used to link the profile to the resources as well as to the activities. More specifically, as shown in Fig. 6, for each activity of the BPMN model a specific distribution profile was created. All the profiles were linked to the corresponding activities through labels, and each resource working on the activity was called “based on the label”. In this way it was possible to build the resource perspective for the BPMN diagram imported in the DES model.

The evaluation of the Resources distribution profile in the DES model resulted for each run of the Trial in the so-called Resources event log (see Fig. 7, where the red-rounded area represents the results of one run of the Trial).

Also in this case, the spreadsheet was exportable as CSV file as allowed input of the ‘.xes’ format in Disco tool, to be in turn exported in ProM6. The deployment of the ‘Replay a Log on Petri-net for Conformance analysis’ plug-in was then needed to confront for each activity the IMi-related Petri Net with the Simulation-related event log, in order to detect the presence of errors in the manual transport of data in Simul8. Not enough matching points were found between the ProM6 data and the DES model data in this case, so the formers could only be used to somehow oversee the organizational perspective of the process in the DES model.

5. Concluding remarks

In this study a novel approach was described to figure out a DES model using Process Mining techniques, for a specific hospital-related process, trying to overcome issues related to: treatment of the data provided from the organization; building time of a DES tool; gaining knowledge of the meaningful process attributes; validation of the resulting model [49].

The main idea of the project – that was to adapt the process mining tool, ProM6, with the commercial Simulation tool, Simul8 –

	A	B	C	D	E	F
1	eerste pkl, bezoek~administratief tarief	Resource 14	98,19731	1,80269	0	
2	Dagverpleging~zwaar	Resource 2	94,98214	5,01786	3,39222	
3	1e consult algemeen	Resource 19	98,8731	1,1269	0,66806	
4	eerste pkl, bezoek~administratief tarief	Resource 10	98,89178	1,10822	1,01427	
5	eerste pkl, bezoek~administratief tarief	Resource 2	96,40908	3,59092	3,32445	
6	eerste pkl, bezoek~administratief tarief	Resource 15	97,07979	2,92021	2,60367	
7	eerste pkl, bezoek~administratief tarief	Resource 3	97,61892	2,38108	2,1855	
8	Dagverpleging~zwaar	Resource 2	95,53206	4,46794	4,2302	
9	Dagverpleging~zwaar	Resource 21	97,57089	2,42911	2,33105	
10	Dagverpleging~zwaar	Resource 2	95,20936	4,79064	4,79064	
11	eerste pkl, bezoek~administratief tarief	Resource 16	95,13318	4,86682	4,69407	
12	Dagverpleging~zwaar	Resource 21	96,18676	3,81324	3,68767	
13	cornea~corneatopografie	Resource 29	99,52268	0,47732	0,30717	
14	Dagverpleging~zwaar	Resource 2	93,40426	6,59574	6,32739	
15	Dagverpleging~zwaar	Resource 21	95,65128	4,34872	4,34872	
16	eerste pkl, bezoek~administratief tarief	Resource 13	94,10792	5,89208	5,66265	
17	Dagverpleging~zwaar	Resource 13	94,72398	5,27602	5,19965	
18	Dagverpleging~zwaar	Resource 30	97,28304	2,71696	2,61297	
19	Dagverpleging~zwaar	Resource 16	94,97708	5,02292	5,02292	
20	eerste pkl, bezoek~administratief tarief	Resource 7	96,91661	3,08339	2,99935	
21	Dagverpleging~zwaar	Resource 13	94,59019	5,40981	5,40981	
22	eerste pkl, bezoek~administratief tarief	Resource 1	97,09604	2,90396	2,76923	
23	eerste pkl, bezoek~administratief tarief	Resource 7	96,70459	3,29541	3,15702	

Fig. 7. Resource Log for one run of the Trial.

was in the overall successfully completed. The original concept idea comes from a critical evaluation of both tools: Simul8 is a commercial tool, which does not allow any system changes, but the user can import, export, connect with Excel, save the work, and many other high level features; ProM6 is a platform that makes possible to build different projects related to the PM techniques in Java source; this allows in many case to export, and to import 'xes' format files. This project shows how the adaptation is possible, implying the deployment of other tools as well, i.e. Disco and Excel, and the steps that have been carried out to get to the final result.

Although more studies are required to find more solid evidence as to the usefulness of PM techniques for determining the DES models construction, the present work provides a robust starting point, as actual links between the two techniques effectiveness were found and demonstrated.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.simpat.2020.102149.

References

[1] C.F. Lindsay, M. Kumar, L. Juleff, L., Operationalising lean in healthcare: the impact of professionalism, Prod. Plann. Control (2019) 1–15.
[2] O. Tamburis, M. Mangia, M. Contenti, G. Mercurio, A. Rossi Mori, The LITIS conceptual framework: measuring eHealth readiness and adoption dynamics across the Healthcare Organizations, Health Technol. (Berl) 2 (2012) 97–112.
[3] J. Hrbcek, V. Simak, Implementation of multi-dimensional model predictive control for critical process with stochastic behavior, Advanced Model Predictive Control, IntechOpen, 2011, pp. 109–124.
[4] C.J. Hamann, J. Loser, L. Reuther, S. Schonberg, J. Wolter, H. Hartig, Quality-assuring scheduling-using stochastic behavior to improve resource utilization, Proceedings of the 22nd IEEE Real-Time Systems Symposium (RTSS), 2001, pp. 119–128.
[5] W.M.P. van der Aalst, A.H.M. ter Hofstede, M. Weske, M., Business process management: a survey, lecture notes in computer science, Proceedings of the 2003 International Conference on Business Process Management, Eindhoven, 26-27 June, 2003, pp. 1–12.
[6] A. Pepino, A. Torri, A. Mazzitelli, O. Tamburis, A simulation model for analyzing the nurse workload in a university hospital ward, 2015 Winter Simulation Conference (WSC), IEEE, 2016, pp. 1367–1378.
[7] J.B. Jun, S.H. Jacobson, J.R. Swisher, Application of discrete-event simulation in health care clinics: a survey, J. Oper. Res. Soc. 50 (1999) 109–123.
[8] C. Baril, V. Gascon, J. Miller, N. Côté, Use of a discrete-event simulation in a Kaizen event: a case study in healthcare, Eur. J. Oper. Res. 249 (2016) 327–339.
[9] M.J. Glover, E. Jones, K.L. Masconi, M.J. Sweeting, S.G. ThompsonSWAN Collaborators, Discrete event simulation for decision modeling in health care: lessons from abdominal aortic aneurysm screening, Med. Decis. Mak. 38 (2018) 439–451.
[10] C. Ellis, Information control nets: a mathematical model of office information flow, Proceedings of the Conference on Simulation, Measurement and Modeling of Computer Systems, ACM Press, Boulder, Colorado, 1979, pp. 225–240.
[11] B. Dongen, A. van Medeiros, H. Verbeek, A. Weijters, W. van der Aalst, The ProM framework: a new era in process mining tool support, in: G. Ciardo, P. Darondeau (Eds.), Application and Theory of Petri Nets, Springer-Verlag, Berlin, 2005, pp. 444–454.
[12] W.M. van der Aalst, M. Netjes, H.A. Reijers, Supporting the full BPM life-cycle using process mining and intelligent redesign, in: contemporary issues in database design and information systems development, Igi Global (2007) 100–132.
[13] Á. Rebuge, D.R. Ferreira, Business process analysis in healthcare environments: a methodology based on process mining, Inf. Syst. 37 (2012) 99–116.
[14] A.A. Mitsyuk, I.S. Shugurov, A.A. Kalenkova, W.M. van der Aalst, Generating event logs for high-level process models, Simul. Model. Pract. Theory 74 (2017) 1–16.
[15] W.M. van der Aalst, Process mining and simulation: a match made in heaven!, Proceedings of the 50th Computer Simulation Conference, Society for Computer

Simulation International, 2018.
[16] R.S. Mans, M.H. Schonenberg, M. Song, W.M. van der Aalst, P.J. Bakker, Application of process mining in healthcare–a case study in a dutch hospital, International Joint Conference on Biomedical Engineering Systems and Technologies, Springer, Berlin, Heidelberg, 2008, pp. 425–438.
[17] E. Rojas, J. Munoz-Gama, M. Sepúlveda, D. Capurro, Process mining in healthcare: a literature review, J. Biomed. Inform. 61 (2016) 224–236.
[18] E. Batista, A. Solanas, Process mining in healthcare: a systematic review, 9th International Conference on Information, Intelligence, Systems and Applications (IISA), IEEE, 2018, pp. 1–6.
[19] O.A. Johnson, T.B. Dhafari, A. Kurniati, E. Fox, Rojas, The clearpath method for care pathway process mining and simulation, International Conference on Business Process Management, Springer, Cham, 2018, pp. 239–250.
[20] W. van der Aalst, A. Adriansyah, A.K.A. De Medeiros, F. Arcieri, T. Baier, T. Blickle, et al., Process mining manifesto, International Conference on Business Process Management, Springer, Berlin, Heidelberg, 2011, pp. 169–194.
[21] M.L. van Eck, X. Lu, S.J. Leemans, W.M. van der Aalst, PM2: a process mining project methodology, International Conference on Advanced Information Systems Engineering, Springer, Cham, 2015, pp. 297–313.
[22] I. Bonacci, O. Tamburis, Deploying new perspectives of network organizations for chronic diseases' integrated management, Int. J. Inf. Syst. Serv. Sect. 2 (2010) 13–27.
[23] M. Polato, A. Sperduti, A. Burattin, M. de Leoni, Time and activity sequence prediction of business process instances, Computing 100 (2018) 1005–1031.
[24] W. Zhao, X. Zhao, Process mining from the organizational perspective, Foundations of Intelligent Systems, Springer, Berlin, Heidelberg, 2014, pp. 701–708.
[25] C. Mauro, T. Happle, A. Sunyaev, J.M. Leimeister, H. Krcmar, From medical processes to workflows: modeling of clinical pathways with the unified modeling language, Proceedings of Third International Conference on Health Informatics, 2010, pp. 388–391.
[26] B.F. van Dongen, J. Carmona, T. Chatain, A unified approach for measuring precision and generalization based on anti-alignments, International Conference on Business Process Management, Springer, Cham, 2016, pp. 39–56.
[27] W.M. van der Aalst, Process mining: discovery, Conformance and Enhancement of Business Processes 2 Springer, Heidelberg, 2011.
[28] J.C. Buijs, B.F. Van Dongen, W.M. van Der Aalst, On the role of fitness, precision, generalization and simplicity in process discovery, OTM Confederated International Conferences" On the Move to Meaningful Internet Systems", Springer, Berlin, Heidelberg, 2012, pp. 305–322.
[29] H.M.W. Verbeek, J.C.A.M. Buijs, B.F. Van Dongen, W.M. van der Aalst, Prom 6: the process mining toolkit, Proceedings of BPM Demonstration Track, 615 2010, pp. 34–39.
[30] http://www.simul8.com/.
[31] T.R. Wallin, M. Hinckley, C. Nilson, R.J. Olson, A clinical comparison of single-piece and three-piece truncated hydrophobic acrylic intraocular lenses, Am. J. Ophthalmol. 136 (2003) 614–619.
[32] H.W. Roberts, K.W. Vijay, D.L. Sullivan, P. Hidzheva, D.I. Detesan, B.S. Heemraz, J.M. Sparrow, D.P.S. O'Brar, A randomized controlled trial comparing femtosecond laser–assisted cataract surgery versus conventional phacoemulsification surgery, J. Cataract Refract. Surg. 45 (2019) 11–20.
[33] M.M. Overduin, Exploration of the Link Between the Execution of a Clinical Process and Its Effectiveness Using Process Mining Techniques, Technische Universiteit Eindhoven, 2013.
[34] G. Sedrakyan, J. De Weerdt, M. Snoeck, Process-mining enabled feedback: tell me what I did wrong vs. tell me how to do it right, Comput. Hum. Behav. 57 (2016) 352–376.
[35] S.J. Leemans, D. Fahland, W.M. van der Aalst, Discovering block-structured process models from event logs-a constructive approach, International Conference on Applications and Theory of Petri Nets and Concurrency, Springer, Berlin, Heidelberg, 2013, pp. 311–329.
[36] K. Aberer, E. Damiani, T. Dillon, Data-Driven Process Discovery and Analysis 116 Springer, 2012.
[37] N. Tax, N. Sidorova, W.M. van der Aalst, R. Haakma, Local process model discovery: bringing Petri nets to the pattern mining world, International Conference on Applications and Theory of Petri Nets and Concurrency, Springer, Cham, 2018, pp. 374–384.
[38] A.O. García, C.D. Lima, I.L. Herrera, Detection of Bottlenecks in Hospital Processes from the XAVIA HIS System using Process Mining, IEEE Latin Am. Trans. 17 (2019) 18–25.
[39] F. Mannhardt, M. De Leoni, H.A. Reijers, W.M. van der Aalst, Balanced multi-perspective checking of process conformance, Computing 98 (2016) 407–437.
[40] T. Allweyer, BPMN 2.0: Introduction to the Standard For Business Process Modelling, BoD–Books on Demand, 2016.
[41] N. Lohmann, M. Song, P. Wohed P (Eds.), Business Process Management Workshops, BPM 2013 International Workshops. Beijing, China, August 26, Revised Papers, 171 Springer, 2013.
[42] I. Yakimov, A. Kirpichnikov, V. Mokshin, Z. Yakhina, R. Gainullin, The comparison of structured modeling and simulation modeling of queueing systems, International Conference on Information Technologies and Mathematical Modelling, Springer, Cham, 2017, pp. 256–267.
[43] S. Bernardi, J.I. Requeno, C. Joubert, A. Romeu, A systematic approach for performance evaluation using process mining: the POSIDONIA operations case study, Proceedings of the 2nd International Workshop on Quality-Aware DevOps, ACM, 2016, pp. 24–29.
[44] H. Nguyen, M. Dumas, A.H. ter Hofstede, M. La Rosa, F.M. Maggi, Business process performance mining with staged process flows, International Conference on Advanced Information Systems Engineering, Springer, Cham, 2016, pp. 167–185.
[45] H. Hassani, E. Silva, A Kolmogorov-Smirnov based test for comparing the predictive accuracy of two sets of forecasts, Econometrics 3 (2015) 590–609.
[46] A. De Myttenaere, B. Golden, B. Le Grand, F. Rossi, Mean absolute percentage error for regression models, Neurocomputing 192 (2016) 38–48.
[47] E. Ruschel, E.A.P. Santos, E.D.F.R. Loures, Establishment of maintenance inspection intervals: an application of process mining techniques in manufacturing, J. Intell. Manuf. 31 (2020) 53–72.
[48] L.E. Chaby, M.J. Sheriff, A.M. Hirrlinger, V.A. Braithwaite, Can we understand how developmental stress enhances performance under future threat with the Yerkes-Dodson law? Commun. Integr. Biol. 8 (2015) e1029689.
[49] O. Tamburis, Bridging the gap between Process Mining and DES modeling in the Healthcare domain, 7th IEEE International Conference on E-Health and Bioengineering, Iasi, Romania, 2019.