AUTOMATED GENERATION OF PATIENT POPULATION FOR DISCRETE-EVENT SIMULATION USING PROCESS MINING

Jules Le Lay
Julia Neveu
Benjamin Dalmas
Vincent Augusto

Mines Saint-Etienne, Univ Clermont Auvergne, INP Clermont Auvergne, CNRS, UMR 6158, LIMOS
Centre CIS, F - 42023 Saint-Etienne FRANCE
{jules.le-lay, benjamin.dalmas, augusto}@emse.fr

ABSTRACT

Process mining is increasingly used to discover and analyze health care processes. It is especially powerful in the study and improvement of patient clinical pathways. Combining process mining results and discrete-event simulation is an interesting approach to discover, represent and assess clinical pathways and improve healthcare organizations. The objective of this work is to develop a framework to automate such studies from the data preprocessing stage to use in simulations. We describe the use of Python and the PM4PY package for formatting data and discovering processes. A generic discrete-event simulation model is developed to serve as a base for analyzing and improving the patient flow in a healthcare center. This type of framework enriches the classical simulation model with synthetic pathways based on real patients and should facilitate accessing aggregated patient data and transposing studies on third-party datasets.

Keywords: Process Mining, Automation, Process Discovery.

1 INTRODUCTION

1.1 Context

This research addresses the need for accessing a large quantity of real patient pathways to establish accurate simulation models. Medico-administrative databases record *all* pathways occurring in healthcare centers. However, some instances, like unique pathways caused by exceptional circumstances, need to be removed so that the model can focus on representative patients.

Since 2016, the General data protection regulation (GDPR), relating to the protection of individuals with regard to the processing of personal data and the free movement of such data, has changed the organization of health systems. Health data are described by GDPR as, "personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status" (European Commission 2016), (article 4).

Macroscopic simulation models allow us to review the stays of patients in independent units and help to improve organization, identify bottlenecks and plan changes in a hospital. This research describes the automatic generation of population information for these types of simulation models.

1.2 Related Literature

Current research has investigated the creation of a complete framework to automate the use of medico-administrative data in simulations. Here, we explore the related literature from different areas of the framework.

Process discovery algorithms map business processes using the data generated by their execution (Aalst 2014). All the steps, or *events*, of an instance form a *trace*. Traces are grouped into collections called *event logs*. Discovery algorithms translate event logs into comprehensible business models, such as Petri nets. Process mining (PM) is extensively used in health care studies and clinical pathway discovery. Of 172 studies, Erdogan et al. reported 156 studies that apply some sort of process discovery methodology, but only 5 involve multiple department pathways (Erdogan and Tarhan 2018). Another review by Rojas et al. reported that 60% of the 74 papers studied investigate healthcare activities from a control-flow perspective (Rojas et al. 2016). The mapping of clinical pathways for multimorbid patients (patients with several conditions at the same time) has been studied for several purposes. Zhang et al. used multimorbid patients medical records to identify the most used services (Zhang and Chen 2012), while Aali et al. was interested in the visualization of the evolution of multimorbidity through a patient's diagnoses over time (Aali, Mannhardt, and Toussaint 2022).

Discrete-event simulation (DES) is an operation research technique that relies on a stochastic modeling approach. In a recent review, (Vázquez-Serrano, Peimbert-García, and Cárdenas-Barrón 2021) identified 231 papers using DES in healthcare and highlighted a continued rise of publications in the last decade. Papers attempting to model healthcare operations tended to focus on one particular department, such as the Emergency department (ED) (Ben-Tovim et al. 2016) or Intensive care unit (ICU) (Busby and Carter 2017), and its interactions with the rest of the hospital. Here, we are interested in *setting-specific simulation models*, as described by (Fletcher and Worthington 2009), in which a generic simulation layout that uses input data to ensure the representation of a specific healthcare center is built. DES is particularly suitable to evaluate the economic aspects of processes. (Soto-Gordoa et al. 2017) built a model to quantify the cost of care of multimorbid patients following an integrated care pathway using Arena®Rockwell software.

(Maruster and van Beest 2009) proposed a simple 3-step approach to combine PM and DES for process improvement: (i) identify performance issues, (ii) map, modify and assess the existing process through simulation and (iii) evaluate the evolution of performance with the redesign case. This research is extended in (Aguirre, Parra, and Alvarado 2013), who added an extra layer of analysis by using data mining and root-cause analysis to improve the process. (Arnolds and Gartner 2018) and (Halawa, Chalil Madathil, and Khasawneh 2021) both used a combination of PM, simulation and optimization to determine the layout of a healthcare center facility. In both cases, PM was used to discover the pathway of patients between different locations and a simulation was used to evaluate the solutions generated by the optimization algorithms.

Frameworks using separate PM and simulation models have been developed recently. (Abohamad, Ramy, and Arisha 2017) combined an ED pathway analysis using the fuzzy miner algorithm with an AnyLogic® DES model. However, the results of the PM analysis were not automatically integrated into the simulation. (Camargo, Dumas, and González-Rojas 2020) proposed a complete framework to discover a process as a business process model network. Nonconforming traces were filtered from a log, and important parameters for the simulation (interarrival times or activity processing times) are simultaneously estimated.

1.3 Objective

In this paper, we propose an automatic framework that generates a population to feed the generic simulation model of a healthcare center and allows modelers to use their own simulated data, which has characteristics similar to those of actual patients. This framework uses historical clinical pathway data stored in a medico-administrative database, filters and maps the clinical pathway of patients in a healthcare center, and generates and formats pathways for the simulation model. For this study, we use the data of multimorbid patients hospitalized in Centre Hospitalier Universitaire de Saint-Étienne (CHUSE) in 2017. The output population is prepared for an AnyLogic® DES model that is adapted from previous work led by our research group.

The remainder of this paper is organized as follows: Section 2 presents the current problems and describes the available data. The developed methodology is detailed in section 3. We describe the population used to test the framework and the numerical results obtained in section 4. Section 5 summarizes the main findings of this research, exposes the limitations of the study and lists opportunities to explore for future research.

2 CURRENT PROBLEMS

Our primary objective is to automatically populate a simulation model with filtered data. To achieve this goal, the proposed methodology needs to focus on two research problems. First, extracting a process map using process mining methods from the history of patient pathways. Numerous PM algorithms exist to generate process maps from event logs. We integrated these methods into our framework to accurately represent the hospitalization process. Secondly, we generate pathways using this process map to respect the stay characteristics of the original population. In particular, the most prevalent pathways and their proportion in the dataset need to be respected, and pathways that are not present in the original dataset need to be avoided. A series of filters must be tested to exclude most unusual pathways and include the most representative ones.

Generating a population from aggregated data avoids the use of personal data. Using this framework in hospitals would facilitate access to essential data (i.e., the most represented pathways and Length of stay (LoS) distribution in medical units) without disclosing personal data. This opportunity, particularly with the implementation of the new gdpr-related provisions, could facilitate the implementation of patient flow studies in hospitals.

The contribution of this article is twofold: (1) to establish a methodology generating a synthetic pool of patients from personal medico-administrative data using process mining and (2) to feed these data into a generic DES for health care process evaluation, allowing the automation of process studies.

3 METHODS

PM is a data mining technique allowing the better understanding and management of processes. The different applications identified are the discovery, analysis of conformance and identification of optimization potentials (Aalst 2014). Therefore, data mining is a bridge that links data science techniques and process sciences.

This section details the proposed framework and its constitutive elements. The complete workflow of this process is presented in Figure 1 and contains three distinct steps. First, process discovery is applied to patient data. Next, the output population is generated using a pathway generation module. Finally, the output population is passed on to a DES simulation.

3.1 Input Data

Medico-administrative databases contain essential information for analyzing the efficiency of patient pathways. Such data include administrative information (such as sex and age) and medical information (Ta-

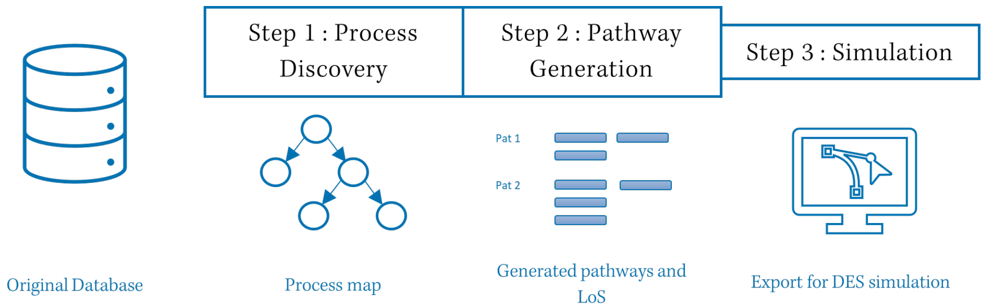


Figure 1: Schematics of the presented process from discovery to simulation.

ble 1). This type of data is considered sensible information. The list of variables accessed and their use in the present study is presented in Table 1.

Table 1: Variable description.

	Variable	Comment
Administrative information	Anonymous identifier	Linking stays
	Age	(no use in the present study)
	Sex	(no use in the present study)
	Residence geographical zone	(no use in the present study)
Medical information	Clinical origin	Analysis and modeling of care pathway
	Visited Medical Units	Analysis, modeling and simulation
	LoS in each unit	Analysis, modeling and simulation
	Admission date	Simulation (arrival law)
	LoS in the hospital	Analysis, modeling and simulation

We only kept information related to the patient pathway. The occupation rate of each medical unit was calculated using the number of occupied beds every day and the theoretical number of beds per service in 2018 (additional data provided by the hospital). Each patient is modeled as an agent associated with the following parameters:

- the anonymous identifier of the patient's hospitalization;
- the pathway of the patient, given as a sequence of all medical units visited by the patient;
- the admission and discharge dates of the patient in each medical unit that compose the sequence; and
- the occupation rate of each medical unit on the admission day in said unit.

An illustrative example is provided in Table 2. It displays two patients whose pathways are composed of 3 and 4 steps, respectively. **Start** and **End** steps were added for each stay, and the date of admission is converted to allow for these steps to be virtually instantaneous.

3.2 Process Discovery

All the described data were imported and processed using the *PM4PY* library (Berti, van Zelst, and Aalst 2019) in Python. Two discovery algorithms were tested to establish the process map, resulting in two different representations: Direct follow graphs (DFGs) and Petri net (PN). They were generated using the *fuzzy*

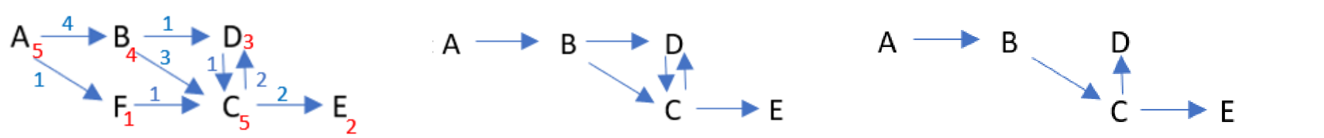
Table 2: Example of event log after pre-processing.

Id	MU	Admission date	Current occupation
Case 1	**Start**	**2017.04.12 00:00:00**	**NA**
	Internal Medicine	2017.04.12 00:00:01	76 %
	Cardiology	2017.04.15 00:00:00	112 % [1]
	Neurology	2017.04.26 00:00:00	84 %
	End	**2017.05.02 00:00:01**	**NA**
Case 2	**Start**	**2017.09.03 00:00:00**	**NA**
	Post-ED unit	2017.09.03 00:00:01	85 %
	Clinical gerontology	2017.09.25 00:00:00	92 %
	Cardiology	2017.09.26 00:00:00	87 %
	Clinical gerontology	2017.10.02 00:00:00	95 %
	End	**2017.10.05 00:00:01**	**NA**

[1] In some circumstances, the number of beds in one unit can be temporarily increased

miner algorithm introduced by (Günther and van der Aalst 2007) and the *inductive miner* algorithm introduced by (Leemans, Fahland, and Aalst 2014). These formats allow for easy export and trace generations with the *PM4PY* library, and these algorithms were chosen because of their preponderance in the literature.

We filter the log using two distinct filters: global and local. The global filter concerns the most represented variants (i.e. a unique sequence of activities) in the log. It allows the removal of the most infrequent pathways from the equation, those who apply to unique pathologies and circumstances and that are mostly unpredictable. Local filters simplify representations (and more precisely DFGs graphs) by removing the least represented activities or transitions from the graph, as shown in Figure 2 on the following set of variants:{ $[A,B,C,D]$; $[A,B,D,C]$; $[A,B,C,E]$; $[A,F,C,E]$; $[A,B,C,D]$ }



(a) Without filter (b) Activity filter (c) Path filter

Figure 2: Example of local filters

We tested the following values for the local filter(percentage of paths we keep): [1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1], combined with global filters of 100%, 95%, 90% or 85%.

3.3 Generating Pathways

Pathway generation is performed in two steps. First, we generated the sequence of visited medical units, and second, we generate the LoS for each step. Both the PN and DFGs models were tested to generate a new population using PM4PY, although the method is different for the two models. The DFGs graph generation is static, i.e. a given model will only generate one set of pathways. We generatec output logs considering the 10, 20, 50, 100 and 150 most frequent pathways. This method can also account for loops, allowing patients to go back and forth between two units. We allowed the generation of pathways with the same activity occurring 2, 5, 7 and 9 times. The PN model allows for dynamically generating logs from a model. We generated logs of 100, 500, 1000 and 5000 pathways, with traces of length 4, 6, 8, 10, 12 or 14 (the maximum number of activities observed). Loops can occur in the pathway but cannot be controlled.

In addition to the traditional metrics used in PM, we develop the 4 following indicators:

- *Variant proportion:* ratio between the number of variants generated and the number of variants observed in the input data;
- *Overlap:* number of variants common to both input and output logs;
- *Under generation:* proportion of variants of the input log not generated by the model; and
- *Overgeneration:* proportion of variants of the output log not present in the input data.

We want to maximize the overlap, which shows how well we recreate the original data, and obtain a variant proportion close to one. In the similar approach, we want to minimize the over- and under-generation figures.

We used the available data to infer LoS distributions and draw a LoS value for each step of each pathway. Two possibilities were tested: (i) LoS distributions were calculated using the LoS history of each unit. (ii) calculate LoS distributions for each pair of units (A, B). This second option was studied to add more precision to the model in case of unit crowdings. Twelve different distributions were tested, and the best fitting distribution was kept for each unit or pair of units. The longest generated stays were limites 1.2 times the maximum duration observed in the data set to remain consistent with the initial observations.

In the simulation, we tracked the final LoS as the sum of a waiting period and a care period to check if it was consistant with the initial hospitalization LoS. We used the same methodology as previously and aggregated the LoS of the patients for each unit. We tested the two following hypotheses: (i) the waiting time in the current unit depends on whether the next unit is crowded, and (ii) the waiting time in the current unit depends on what the next service is and on whether it is crowded. We created both subsets of data and calculated the distribution for both cases.

Generated patients, their pathways, and the calculated LoS distributions are automatically stored and exported in a normalized excel file for the simulation.

3.4 Simulation

A simple macroscopic simulation model of a generic hospital was developed using AnyLogic®software. Patients created in this model are randomly provided a pathway and a corresponding LoS sequence from the log generated previously. The patients' pathway are passed in the simulation as parameters. The only resource we considered is the number of beds per unit (staff utilization is not modeled). They were modeled using ResourcePool objects and dynamically chosen in seize blocks depending on the unit requested in the pathway. The model was thoroughly validated using different datasets to ensure its proper function.

4 RESULTS

The data used for the whole project relates to multimorbid patients hospitalized at the hospital of Saint Étienne in 2017. To identify patients in the database, multimorbidity was defined as having one diagnosis in at least three different chapters of the International Statistical Classification of Diseases and Related Health Problems 10th Revision (IDC-10) classification. A lot of diagnoses per patients are listed in medical units, and a vast majority of patients have diagnoses in 2 or more chapters. This first filter was made to exclude non multimorbid patients and was performed by the hospital administration. Data were provided by the CHUSE under Commission Nationale de l'Informatique et des Libertés (CNIL) authorization number 919300.

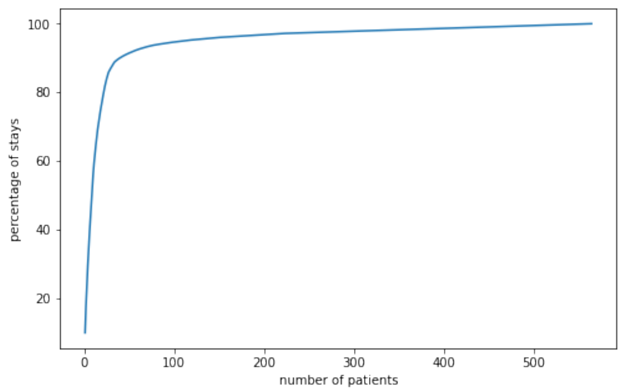


Figure 3: Percentage of patients by number of variants.

4.1 Process Discovery

Figure 3 shows the cumulative percentage of patients per variant. The first variant accounts for more than 23% of patients, and 90% of patients leave with only 21 variants. The DFG and PN graphs built with the *fuzzy miner* and *inductive miner* algorithms and an 85% filter (this filter value was chosen for readability reasons) on log variants are displayed in Figure 4 and Figure 5, respectively.

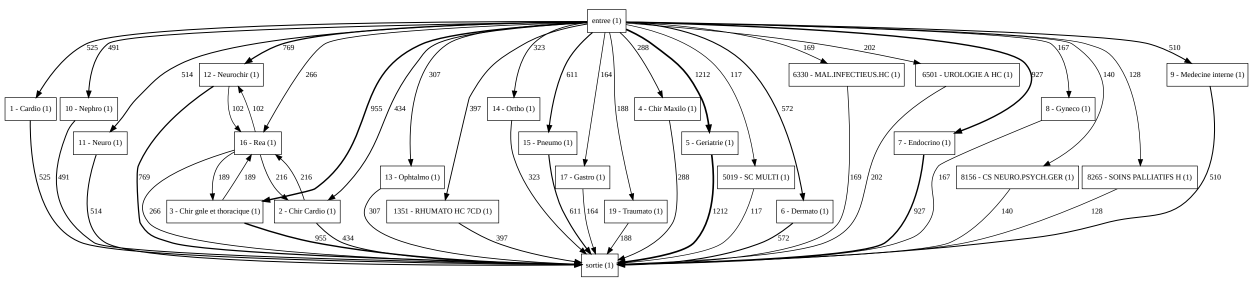


Figure 4: Graph obtained with an 85% global filter on log variants.

These representations reveal three clusters of medical units. The first cluster groups units with patients who come for "simple" stays, with only one unit visited, such as the endocrinology, gastroenterology, or gerontology units. The second cluster is organized around post-emergency services in more complex pathways. The last cluster is composed of medical units with high resource utilization, such as glsICUs and surgery.

The fitness, precision, generalization and simplicity when using the global filters are displayed in Figure 6 for the two studied models. Fitness indicates how the model allows us to replay paths of the event log, simplicity allows us to select simpler representations, precision measures the ability to discriminate traces that are not in the original log, and generalization allows us to eliminate overly specific models. The two values of global filters that seem to meet the evaluation criteria are the 95% and 90% filters. These values allow the model to achieve a good balance between generalization and precision. For the local filter, keeping the 90% most frequent paths and activities on the log filtered to keep the 95% most frequent pathways gives satisfying results. The output fitness, accuracy and generalization values are close to 1, indicating good overall precision while keeping the simplicity metric at a satisfying level.

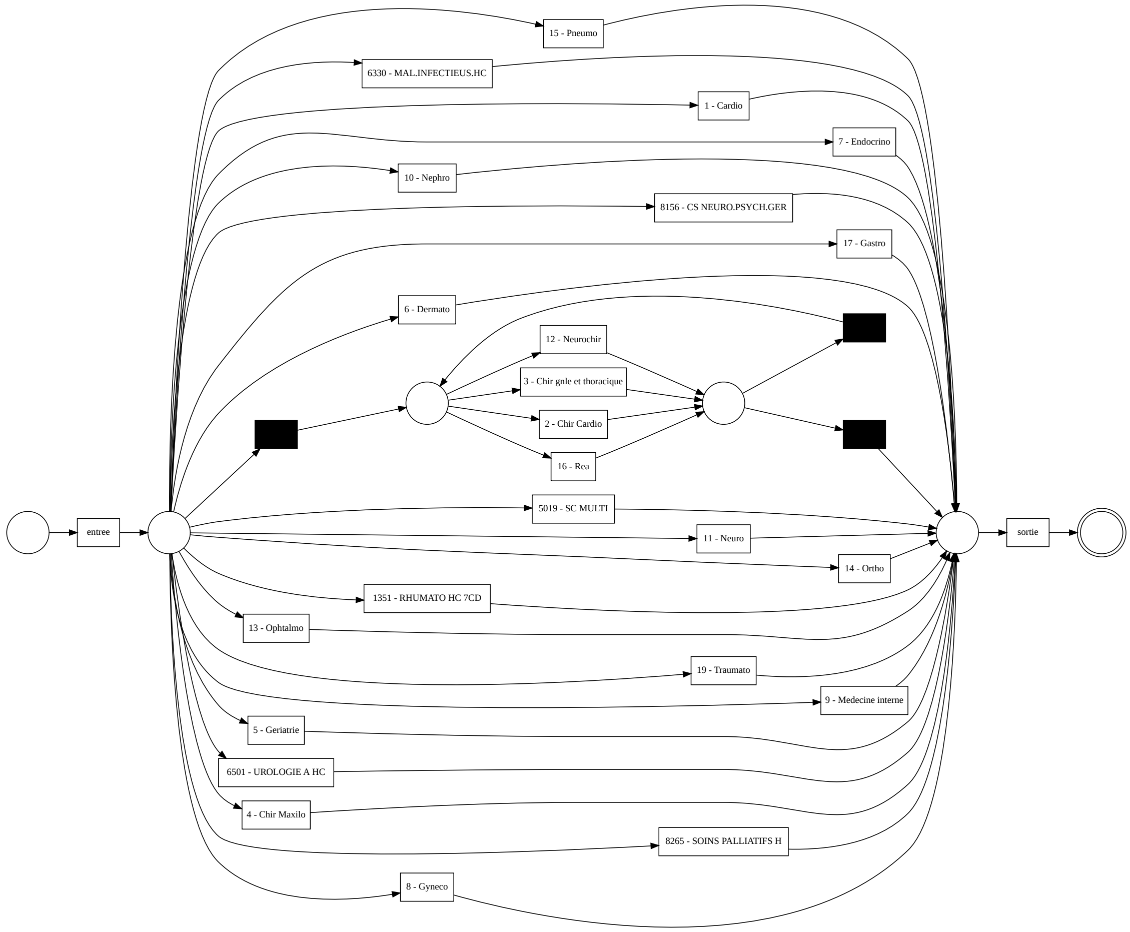


Figure 5: Petri net obtained with an 85% global filter on log variants.

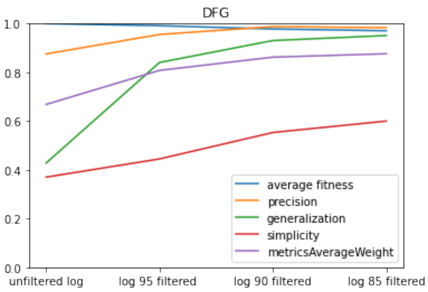
4.2 Log Generation

The results for the DFG and Petri net models, with the global filter set to 95% are shown in Table 8. Metrics are grouped into 4 cells as follows: variant proportion in top-left cells, overlap in top-right cells, undergeneration in bottom-left cells and overgeneration in bottom-right cells. The initial time values and the calculated distribution for the cardiology unit are displayed in Figure 9.

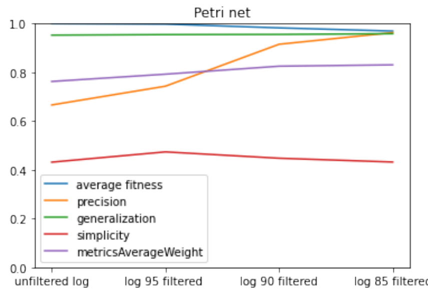
The distributions generated for pairs of units were rarely significant because of the insufficient number of observations for many of the paths considered. For statistically significant distributions, we obtained a shorter average LoS when the next unit was crowded. These counterintuitive results and the absence of results for some pairs of units prompted us to use distributions of units considered independently.

5 CONCLUSIONS AND PERSPECTIVES

The present paper develops an automated framework to use generates populations of patients in simulation. The proposed framework generates process mining models from patients electronic records and generates a population mimicking patients pathways and LoS of the initial population. The resulting log is described using a set of metrics that can be used to fine tune the parameters that define the process model and population generation. The output population is automatically formatted to be ready to use in a generic DES. This simulation model accurately replays clinical pathways with realistic LoS (average obtained of 8.5 days, sim-



(a) Using DFG



(b) Using PN

Figure 6: Global filter effect on metrics.

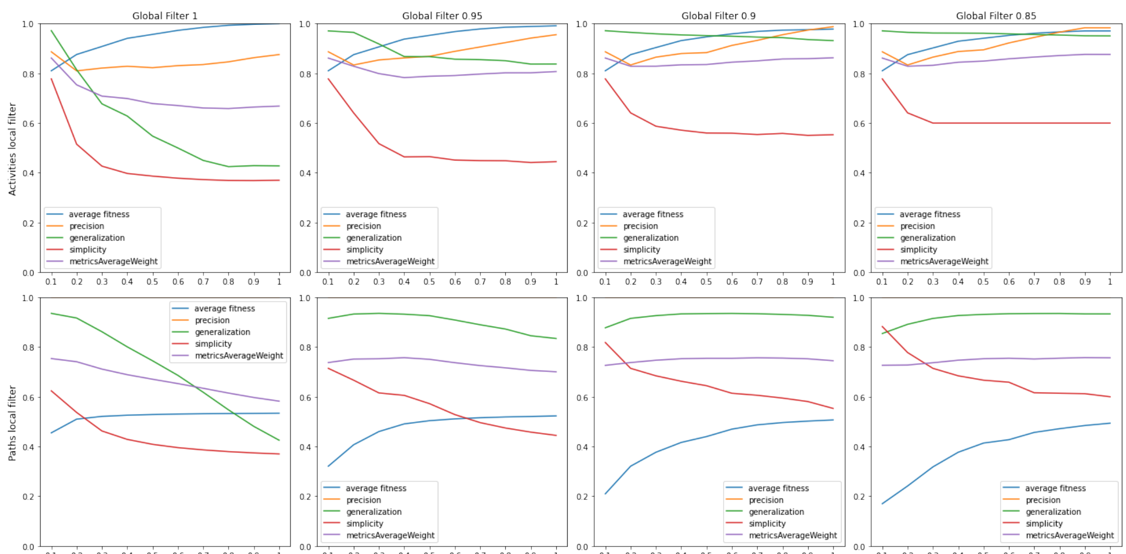


Figure 7: Evaluation of the DFG model based on the global filter on activities and local filters on the paths and activities of the DFG model.

ilar to the initial population) and allows modelers to account for phenomena that cannot be detected when using only process mining, such as the crowding of medical units.

This framework has several limitations. Although accurate when considering units, the LoS generation module cannot be used to model more precise transitions between units or waiting time, mainly because of the low number of patients that visited. We believe that considering only multimorbid patients may bias this dataset by over- or under-representing some pathways. Further investigations should be performed to remedy this problem and increase the possibilities of the framework. The generation module only allows for the generation of pathways and LoS. Other important features, such as age or diagnosis, are not considered, making it difficult to state the credibility of the pathway.

Improvements are also made to the simulation model. The admissions rates in the hospital were estimated using multimorbid patients' records and were corrected to account for non-multimorbid patients. In addition, the macroscopic layout of the simulation model and the fact that it only accounts for bed resources make it hard to use for microcosting analysis that is usually carried out with DES. The consumption of medical resources or patient admission constraints related to staff obligations (housekeeping or patient accompani-

Filter	Nb variants Max	Max activity occurrence 2		Max activity occurrence 5		Max activity occurrence 7		Max activity occurrence 9	
0,95	150	0,33	0,71	0,33	0,71	0,33	0,71	0,33	0,71
0,95	150	1,02	0,29	1,02	0,29	1,02	0,29	1,02	0,29
0,95	100	0,22	0,83	0,22	0,83	0,22	0,83	0,22	0,83
0,95	100	1,07	0,17	1,07	0,17	1,07	0,17	1,07	0,17
0,95	50	0,11	0,90	0,11	0,90	0,11	0,90	0,11	0,90
0,95	50	1,16	0,10	1,16	0,10	1,16	0,10	1,16	0,10
0,95	20	0,04	1,00	0,04	1,00	0,04	1,00	0,04	1,00
0,95	20	1,21	0,00	1,21	0,00	1,21	0,00	1,21	0,00
0,95	10	0,02	1,00	0,02	1,00	0,02	1,00	0,02	1,00
0,95	10	1,24	0,00	1,24	0,00	1,24	0,00	1,24	0,00

(a) Metrics obtained for the DFG model.

Filter	Nb variants Max	Max length 4		Max length 6		Max length 8		Max length 10		Max length 12	
0,95	5000	1,80	0,11	2,18	0,08	2,08	0,09	2,11	0,10	2,05	0,09
0,95	5000	1,07	0,89	1,08	0,92	1,07	0,91	1,06	0,90	1,07	0,91
0,95	1000	0,55	0,19	0,56	0,20	0,51	0,20	0,57	0,17	0,59	0,20
0,95	1000	1,16	0,81	1,15	0,80	1,16	0,80	1,16	0,83	1,14	0,80
0,95	500	0,32	0,22	0,32	0,19	0,33	0,21	0,34	0,25	0,32	0,24
0,95	500	1,19	0,78	1,20	0,81	1,19	0,79	1,17	0,75	1,18	0,76
0,95	100	0,10	0,48	0,09	0,46	0,10	0,43	0,09	0,34	0,09	0,31
0,95	100	1,21	0,52	1,22	0,54	1,22	0,57	1,23	0,66	1,23	0,69

(b) Metrics obtained for the PN model (results for variant length 14 are not presented to improve the readability).

Figure 8: Evaluation metrics results table for the DFG and PN model with a 95% global filter.

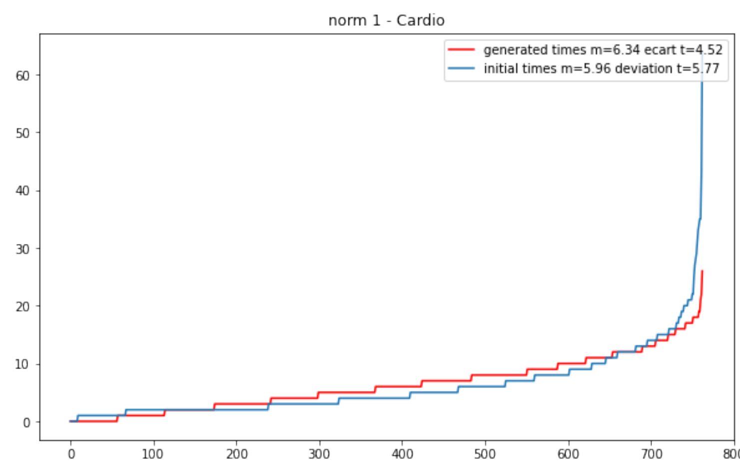


Figure 9: Example of the fitted distribution.

ment) cannot be properly modeled currently. Here, we used data aggregated at the unit level, and accessing analyses such as those performed in (Arias et al. 2020) would allow us to generate patient pathways with more details and perform cost analysis accounting for a large number of resources.

However, the model's ability to accurately replay pathways and to queue patients makes it a reliable tool for strategic decision making, for instance, to estimate the effects of a bed shortage in a unit or the change of care pathway for a particular type of patient.

REFERENCES

Aali, M. N., F. Mannhardt, and P. J. Toussaint. 2022. "Discovering Care Pathways for Multi-morbid Patients Using Event Graphs". In *Process Mining Workshops*, edited by J. Munoz-Gama and X. Lu, pp. 352–364. Cham, Springer International Publishing.

Aalst, W. v. d. 2014. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer Berlin Heidelberg.

Abohamad, W., A. Ramy, and A. Arisha. 2017. "A hybrid process-mining approach for simulation modeling". In *2017 Winter Simulation Conference (WSC)*, edited by W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, pp. 1527–1538. Piscataway, New Jersey, Institute of Electrical and Electronics Engineers.

Aguirre, S., C. Parra, and J. Alvarado. 2013. "Combination of Process Mining and Simulation Techniques for Business Process Redesign: A Methodological Approach". In *Data-Driven Process Discovery and Analysis*, edited by W. van der Aalst, J. Mylopoulos, M. Rosemann, M. J. Shaw, C. Szyperski, P. Cudre-Mauroux, P. Ceravolo, and D. Gašević, Volume 162, pp. 24–43. Berlin, Heidelberg, Springer Berlin Heidelberg. Series Title: Lecture Notes in Business Information Processing.

Arias, M., E. Rojas, S. Aguirre, F. Cornejo, J. Munoz-Gama, M. Sepúlveda, and D. Capurro. 2020. "Mapping the Patient's Journey in Healthcare through Process Mining". *International Journal of Environmental Research and Public Health* vol. 17, pp. 6586.

Arnolds, I. V., and D. Gartner. 2018. "Improving hospital layout planning through clinical pathway mining". *Annals of Operations Research* vol. 263 (1), pp. 453–477.

Ben-Tovim, D., J. Filar, P. Hakendorf, S. Qin, C. Thompson, and D. Ward. 2016. "Hospital Event Simulation Model: Arrivals to Discharge–Design, development and application". *Simulation Modelling Practice and Theory* vol. 68, pp. 80–94.

Berti, A., S. van Zelst, and W. Aalst. 2019. "Process Mining for Python (PM4Py): Bridging the Gap Between Process- and Data Science". In *Proceedings of the ICPM Demo Track 2019*, edited by A. Burattin, A. Polyvyanyy, and S. van Zelst. Aachen, Germany.

Busby, C. R., and M. W. Carter. 2017. "Data-Driven Generic Discrete Event Simulation Model of Hospital Patient Flow Considering Surge". In *2017 Winter Simulation Conference*, edited by W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, pp. 3006–3017. Piscataway, New Jersey, Institute of Electrical and Electronics Engineers, Inc.

Camargo, M., M. Dumas, and O. González-Rojas. 2020. "Automated discovery of business process simulation models from event logs". *Decision Support Systems* vol. 134, pp. 113284.

Erdogan, T. G., and A. Tarhan. 2018. "Systematic Mapping of Process Mining Studies in Healthcare". *IEEE Access* vol. 6, pp. 24543–24567.

European Commission 2016. "REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)". Technical report, Brussels, Belgium.

Fletcher, A., and D. Worthington. 2009. "What is a 'generic' hospital model?—a comparison of 'generic' and 'specific' hospital models of emergency patient flows". *Health Care Management Science* vol. 12, pp. 374.

Günther, C. W., and W. M. P. van der Aalst. 2007. "Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics". In *Business Process Management*, edited by G. Alonso, P. Dadam, and M. Rosemann, Lecture Notes in Computer Science, pp. 328–343. Berlin, Heidelberg, Springer.

Halawa, F., S. Chalil Madathil, and M. T. Khasawneh. 2021. “Integrated framework of process mining and simulation–optimization for pod structured clinical layout design”. *Expert Systems with Applications* vol. 185, pp. 115696.

Leemans, S., D. Fahland, and W. Aalst. 2014. “Process and Deviation Exploration with Inductive Visual Miner.”. In *BPM Demo Sessions 2014*, edited by L. Limonad and B. Weber, pp. 46. Eindhoven, the Netherlands.

Maruster, L., and N. R. T. P. van Beest. 2009. “Redesigning business processes: a methodology based on simulation and process mining techniques”. *Knowledge and Information Systems* vol. 21, pp. 267.

Rojas, E., J. Munoz-Gama, M. Sepúlveda, and D. Capurro. 2016. “Process mining in healthcare: A literature review”. *Journal of Biomedical Informatics* vol. 61, pp. 224–236.

Soto-Gordoa, M., A. Arrospide, M. Merino Hernández, J. Mora Amengual, A. Fullaondo Zabala, I. Larrañaga, E. de Manuel, and J. Mar. 2017. “Incorporating Budget Impact Analysis in the Implementation of Complex Interventions: A Case of an Integrated Intervention for Multimorbid Patients within the CareWell Study”. *Value in Health* vol. 20, pp. 100–106.

Vázquez-Serrano, J. I., R. E. Peimbert-García, and L. E. Cárdenas-Barrón. 2021. “Discrete-Event Simulation Modeling in Healthcare: A Comprehensive Review”. *International Journal of Environmental Research and Public Health* vol. 18, pp. 12262.

Zhang, X., and S. Chen. 2012. “Pathway identification via process mining for patients with multiple conditions”. In *2012 Institute of Electrical and Electronics Engineers International Conference on Industrial Engineering and Engineering Management*, pp. 1754–1758. Hong Kong, Institute of Electrical and Electronics Engineers.

AUTHOR BIOGRAPHIES

JULES LE LAY is currently a Ph.D. student at the Center for Health Engineering at Mines Saint-Étienne. He received an MSc in engineering from Mines Saint-Étienne in 2017. His research interests include modeling, simulation and optimization of healthcare systems. His email address is jules.le-lay@emse.fr.

JULIA NEVEU is an MSc student in engineering at ESME Sudria. She performed an internship at Mines Saint-Étienne on this project. Her research interests include computer sciences and data analysis.

BENJAMIN DALMAS is an assistant professor in healthcare data science in the Department of Biomedical and Healthcare Engineering. He holds a Ph.D. from Clermont-Auvergne University for his work on the modeling of healthcare pathways. His research interests are the modeling and prediction of patient pathways through the analysis of medical and administrative data using a wide range of techniques, from process and data mining to machine learning. His email address is benjamin.dalmas@emse.fr.

VINCENT AUGUSTO is a professor and a permanent faculty member of Mines Saint-Étienne, Center for Biomedical and Healthcare Engineering. His areas of research comprise performance evaluation, optimization and management of healthcare systems, operational research, health data and medical decision aid (process mining, machine learning). He holds a Ph.D. for his work on the modeling, analysis and control of flows in healthcare systems. He defended his "Habilitation à Diriger des Recherches" on Health Territories Engineering in 2016. He was a visiting professor at CIRRELT (Inter-university Research Center on Business Networks, Logistics and Transport), Laval University, Québec, Canada, in 2009 and 2015. He has been co-responsible for the living lab MedTechLab and director of the Center for Biomedical and Healthcare Engineering of Mines Saint-Étienne since January 2020. His email address is augusto@emse.fr.