sciendo

Process Mining Contributions to Discrete-event Simulation Modelling

Mario Jadrić
University of Split, Faculty of Economics, Business and Tourism, Split, Croatia
Ivana Ninčević Pašalić
University of Split, Faculty of Economics, Business and Tourism, Split, Croatia
Maja Ćukušić
University of Split, Faculty of Economics, Business and Tourism, Split, Croatia

Abstract

Background: Over the last 20 years, process mining has become a vibrant research area due to the advances in data management technologies and techniques and the advent of new process mining tools. Recently, the links between process mining and simulation modelling have become an area of interest. **Objectives:** The objective of the paper was to demonstrate and assess the role of process mining results as an input for discrete-event simulation modelling, using two different datasets, one of which is considered data-poor while the other one data-rich. **Methods/Approach:** Statistical calculations and process maps were prepared and presented based on the event log data from two case studies (smart mobility and higher education) using a process mining tool. Then, the implications of the results across the building blocks (entities, activities, control-flows, and resources) of simulation modelling are discussed. **Results:** Apart from providing a rationale and the framework for simulation that is more efficient modelling based on process mining results, the paper provides contributions in the two case studies by deliberating and identifying potential research topics that could be tackled and supported by the new combined approach. **Conclusions:** Event logs and process mining provide valuable information and techniques that could be a useful input for simulation modelling, especially in the first steps of building discrete-event models, but also for validation purposes.

Keywords: process mining, event log, simulation model, smart mobility

JEL classification: M15
Paper type: Research article

Received: Feb 3, 2020
Accepted: Apr 24, 2020

Acknowledgements: This research is supported by the Croatian Science Foundation [grant number UIP-2017-05-7625].

Citation: Jadrić, M., Ninčević Pašalić, I., Ćukušić, M. (2020), "Process Mining Contributions to Discrete-event Simulation Modelling", Business Systems Research, Vol. 11, No. 2, pp. 51-72.
DOI: 10.2478/bsrj-2020-0015

Introduction

Process mining (PM) is a relatively new scientific domain; as organisations needed to learn more about how their processes function in the real world, concepts and tools related to business process mining emerged (Tiwari et al., 2008). The advent of PM was characterized by different algorithmic approaches concentrating on log data. For example, Agrawal et al. (1998) constructed a process flow graph from execution logs of a workflow application. At the same time, Cook and Wolf (1998) attempted to discover the process models from event log data. Van Der Aalst et al. (2004) affirmed the importance of transactional information systems that can provide appropriate (log) data. PM derived from data mining, referring specifically to the extraction of knowledge from large data sets by identifying patterns within data (Tiwari et al., 2008). As an adaption of data mining, PM is defined as a methodology for obtaining process models from event logs associated with particular processes (Van Der Aalst, 2011).

To achieve the objectives set at the individual and institutional level, organisations focus on improving their business processes. Simulation modelling (SM) has helped in understanding the complexity of business processes, although the main challenge remained to be the development of a realistic simulation model. With technological advances, business processes are increasingly supported by information systems that generate event logs (Abohamad et al., 2017). The logs contain important information about the actual performance of a process and can be analysed using process mining techniques (Van Der Aalst, 2011). Consequently, PM techniques help gain new insights, which can prove to be beneficial as an input in conceptual simulation modelling stages. The output of PM is a process model consisting of a set of activity models and constraints between them (Weske, 2012). Relation and relevance of the model in the context of simulation modelling is the focus of the paper.

The rising number of publications in the Web of Science and Scopus databases for "process mining" AND "simulation" is noteworthy, clearly indicating the emerging research trend in the area studied since 2012. Despite a growing number of studies that relate the benefits of coupling process mining with simulation modelling, the research area still has some challenges that need addressing (Van Der Aalst, 2012; Zakarija et al., 2020). Therefore, this exploratory study aims to address particular scientific challenges noted by the IEEE's Task Force (2011) and to demonstrate the relevance of process mining in specific settings. In consideration of specific challenges identified in PM manifesto (IEEE TFPM, 2011) related to finding/merging/cleaning event data, handling event logs with different characteristics and combining process mining with other types of analysis, the following research questions were formulated in this study:

RQ1: How different sets of event data and event logs affect process mining results?

RQ2: How useful is a combination of process mining with other types of analysis, in particular - how PM results perform as input for discrete-event simulation modelling?

Therefore, the objective of the paper is to demonstrate and assess the potential of using process mining results as an input for discrete-event simulation modelling. The process mining procedure is elaborated using two datasets in two different domains, of which one dataset is considered as data-poor (smart parking) and the other one data-rich (learning management system).

The paper is organised as follows. Chapter 2 presents an overview of key concepts, including PM types, techniques, event logs, and linking PM with discrete-event SM. Chapter 3 outlines and discusses the methodology of the exploratory study, including quantitative and qualitative analyses of cases in two different domains. Chapter 4 discusses the results of the study by comparing it with existing research, explains contributions to the research area, and puts forward implications for the practice. Chapter 5 concludes the study, lists the limitations and suggestions for future research.

Theoretical background

Process mining types and techniques

According to Van Der Aalst et al. (2010), there are three basic types of business process mining: 1) Process discovery where there is no *a priori* model, i.e. a process model is discovered from an event log without prior knowledge of the process itself; 2) Conformance checking where *a priori* model for a process is known and is used for detecting, locating and explaining possible deviations from the standard process (Rovani et al., 2015); and (3) Extension where *a priori* model is known, and it is improved by further data adding a new perspective. Later, Van Der Aalst (2011) replaced the third type of PM with a new term – enhancement, denoting two different types of enhancement: repair (aiming to modify the model to reflect the reality better) and extension (adding a new perspective to the process model). Generally, three perspectives in business process mining are highlighted (Dustdar et al., 2005): 1) process perspective where the aim is to find a suitable representation of all possible paths within the process; 2) organisational perspective, which focuses on people, their roles and their relationships; and 3) case perspective that considers attributes that differentiate one process (case) path from the other. Similarly, Van Der Aalst (2011) distinguishes between four perspectives of PM application: 1) control-flow perspective where the focus is on process structure discovery ("How?"), 2) organisational or resource perspective focusing on resource information ("Who?"), 3) case/data perspective focusing on properties of cases ("What?"), and 4) time perspective related to the timing and frequency of events ("When?").

With regards to business process mining techniques, Turner et al. (2012) list five categories: 1) Transition systems and regions – a technique that mines the models that offer a balance between over- and under-fitting the event log (Van Der Aalst et al., 2010); 2) Clustering techniques - techniques that provide the possibility to combine different process mining approaches in order to mine more challenging event logs, e.g. those containing noise; 3) Heuristic approach – an approach set by Weijters and Van Der Aalst (2001, 2003) who established a set of rules to determine the precedence between tasks and overall task sequences; 4) Evolutionary techniques (such as genetic algorithm by de Mederios et al. (2005) aimed to mine logs containing noise and duplicate tasks); 5) Declarative mining approach suggested by Cattafi et al. (2010) that recognises that the changes in process are possible over time, and revisions of the mined model will take into account new process evidences and deviations. Focusing on PM techniques used for control-flow discovery exclusively, Rojas et al. (2016) emphasised that the most frequently used PM techniques are: 1) Heuristics Miner which successfully deals with noise in event logs thus generating strong business process models (Weijters et al., 2006); 2) Fuzzy Miner which can generate multiple process control models based on fuzzy criteria; and 3) Trace Clustering techniques that can discover simple process models by dividing into partitions (Song et al., 2009).

In their review study, Tiwari et al. (2008) analysed the main techniques used for PM concluding that in 39% of the reviewed papers authors developed custom algorithmic approaches to process mining, while in 10% of papers authors used data mining-based approaches, 6% used soft computing techniques for the process-mining tasks, and in 29% of reviewed papers authors detailed the use of Petri net modelling.

With the growing popularity of the concept, the tools for PM were developed: Disco (by Fluxicon), ARIS Process Performance Manager (by Software AG), Perceptive

Process Mining (by Perceptive Software), Celonis Process Mining (by Celonis GmbH), Process Analyzer (by QPR), Interstage Process Discovery (by Fujitsu), Discovery Analyst (by Stereo LOGIC), and XMAnalyzer (by XMPro) to list the most important ones.

From event data to event logs

Event logs are sets of traces where each trace describes the lifecycle of a particular case (i.e., a process instance) in terms of the activities executed, resources engaged for completing the activity, timestamp, and other information (Van Der Aalst, 2015). A simple example of an event is presented in Table 1.

Table 1
Example of an Event Recorded in an Event Log

Case id	Timestamp	Activity	Resource	Cost
123456	30-12-2019:11.02	register request	John	1000

Source: Authors' example

In this context, the problems of incomplete and raw data often occur, and to deal with this particular issue authors attempted to review and broaden the definition of events. Events are described by references and attributes where references designate a name of an object (for example a person, machine, product), while attributes can denote a name, value, and can contain perspectives like time, age, function, or category (Van Der Aalst, 2015; Wang, 2018). Relevant guidelines for data logging, set by Van Der Aalst (2015) include: 1) precise semantics of reference and variable names as they should have the same meaning for all stakeholders, 2) reference and variable names should be recorded in a structured and managed collection, 3) references should be stable, and independent to different identifiers, 4) attribute values should be precise, 5) occurrence, references or attributes of the events should be clearly marked in cases of uncertainty, 6) events should be ordered, if possible, using the timestamp variable, 7) transactional information about an event should be stored, 8) syntax correctness of the event log should be checked regularly, 9) comparability of event logs needs to be ensured, 10) events should be detailed and not aggregated in the event log which will be an input to the analysis process, and 11) events should not be deleted. Wang (2018) added additional four guidelines as a part of the D2FD (Data to Fuzzy-DEVS) method: 12) privacy should be ensured without losing correlations, 13) event data should be recorded in .csv format or excel files, 14) value names should be simple and clear, 15) the start time and the finish time in the attribute properties are mandatory. The same author complements the list with a general recommendation to 16) order events by instance and then by start date (where instance defines a specific sequence of a case, and a reference is identified as a label of instance).

Creating an event log from raw event data consists of the following steps (Van Der Aalst, 2015): 1) selecting relevant events for the process at hand, 2) correlating events to form process instances, 3) ordering the events using timestamp information, and 4) selecting or computing event attributes based on the raw data (resource, cost, and more). Similarly, Wang (2018) proposed a five-step method to transform the event data to event logs: 1) defining the goal – the first step of the method implies addressing the problem and determining the purpose of performance evaluation, 2) identifying the relationships which can help researchers in finding the analysis results, 3) identifying the values, 4) selecting the process instance, and 5) mapping between event data and event logs.

From process mining to discrete-event simulation modelling

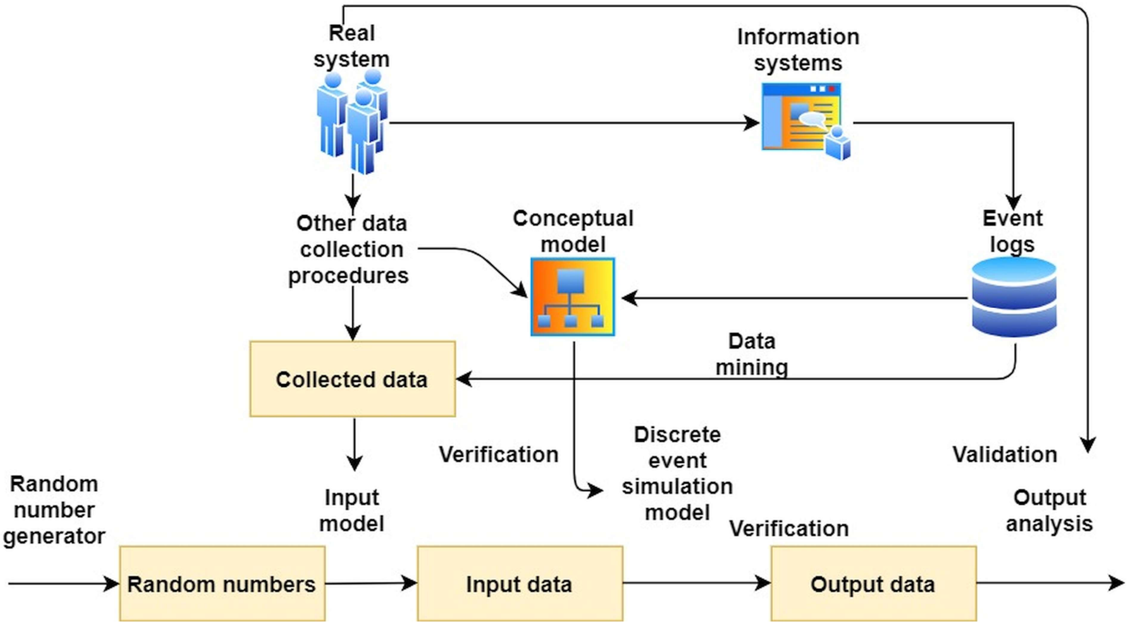
New studies emerged, describing attempts to integrate process mining techniques with simulation modelling. As this is an area to which this paper aims to contribute to, this section provides a brief overview of the work done in that regard.

Martin et al. (2015) presented their conceptual model of how PM techniques can be used to support the construction of simulation models. They used an illustrative case study with a company that provides roadside assistance services to demonstrate that the combination of event log knowledge and traditional information sources (expert interviews) is complementary in the process of constructing a simulation model. Abohamad et al. (2017) presented a framework for integrating process mining techniques into the conceptual modelling phase, i.e. in one of the steps in developing simulation models. The proposed hybrid framework was then tested to determine the performance bottlenecks and explore possible improvement strategies within a hospital's emergency department. In the same application domain - health, Zhou et al. (2014) tried to combine process mining with simulation models to develop an outpatient clinic model. A more general approach was proposed by Rozinat et al. (2009) where the authors developed a simulation model presented as a Coloured Petri Network (CPN) for process analysis and performance evaluation of different models. The authors used a combination of process mining techniques to detect control-flow, data, performance, and process resources based on historical data, and integrated them into the CPN simulation model. Another recent use case was presented by Jadrić (2019), focusing on possible issues and advantages of the detection of student behaviour and processes based on the data from a standard learning management system. He then proposed a conceptual model for using process mining techniques to support the development of discrete-event simulation models in an educational environment.

Discrete-Event Simulation (DES) in this paper denotes multiple computer simulation approaches based on the general idea of modelling a state of a discrete dynamic system which is composed of "state variables" provided by the attributes of "entities" which represent real objects (Wagner et al., 2016). Modelling the system dynamics is done through modelling the events that are reflective of system state changes (ibid). In that context, an event is considered as something that triggers the next set of state variable changes (Liu, 2015). Discrete-event modelling process generally progresses through several steps (illustrated in Figure 1).

A brief outline of the illustrated steps is provided herein: 1) data collection - following standard sampling practices to collect data on the real system in a representative way, 2) random number generation - used to generate a set of random numbers for an input model, 3) determining the input model - to transform a set of random numbers into a set of input data based on the selected statistical distribution; the transformation process is called the generation of random variables, 4) development of conceptual and computational models and their verification, 5) analysis of the output generated by the use/deployment of a computational model based on a set of input data for model validation, and 6) statistical analysis of output data, including, for example, calculating mean, median, standard deviation and confidence intervals, and then experimenting using the simulation model.

Figure 1
Discrete-event modelling process steps



Source: Authors' illustration

To discrete-event modelling process, event logs and process mining can provide valuable information and techniques that could be used for simulation modelling, especially in the first steps of building discrete-event models. The potentials and the extent of the use of event logs in building simulation models are elaborated by Martin et al. (2015) for each of the aggregated simulation model building blocks: entities, activities, control-flow, and resources. Table 2 provides an overview of the role of process mining in simulation modelling by matching aggregated building blocks to relevant process mining features.

Table 2
Overview of PM Research Contributions in Simulation Modelling

Model component	Modelling task	Process mining contribution
Entities	Entity attributes	After the appropriate mapping of case and event attributes to entity attributes is ascertained, logs are used to analyse and assign attribute values to entities.
	Entity type	The authors have not identified examples and research efforts to support entity type modelling using event logs. However, the potential is evident, considering different event log information that is frequently recorded.
	Entity arrival date	Actual entity arrival is critical information as it has a significant influence on the process performance (e.g., the average queue length). This information can usually be extracted from event logs.

Activities	Activity definition	Event logs can be helpful when defining simulation model activities since, by definition, activities record events in the log; however, particular focus should be put on the fact that event logs might register even more detailed information (Baier et al., 2013).
	Activity duration	Due to an extensive array of activity duration determinants (Pospišil & Hruška, 2012), their modelling is complex. However, the duration can be determined from the logs as the difference between the timestamps of start and end events.
	Resource requirements	PM is only useful in this regard if an event log contains resource information.
	Queue discipline	Authors acknowledge there is little research done in analysing topics related to queuing despite the PM potential.
	Queue abandonment condition	Potential to explore the topic is enormous as event log registers the abandonment as well.
	Interruptibility and unexpected interruptions	A combination of outlier analysis and log-based resource schedules can support the identification of interruptible activities.
Control-Flow	Control-flow definition	PM provides the most extensive support for control-flow discovery as the sequential relationship between activities, and gateway type can be discovered by analysing log data.
	Gateway routing logic	Event logs can support modelling of routing logic by analysing circumstances of activity execution such as the probabilities and frequencies of different activity routes.
Resources	Resource roles	Event logs contain resource information that can support resource role identification.
	Resource schedule	Event log analysis for resource scheduling is complicated; however, it can be helpful in some cases.
	Unavailability handling procedure	This aspect can be handled as a part of the research on the discovery of resource schedules.
	Entity handling procedure	An analysis of the activities performed by a resource is required when modelling entity handling procedure using event logs.

Source: Adapted from Martin et al. (2015)

Methodology

This part of the paper describes the methodology used to address research questions set in the introductory section. First, background information is provided based on the literature review for both domains of the case studies (the first one from smart mobility area and the second from the higher education). After that, the rationale for selecting the case studies is presented, followed by a description of PM procedures and analyses where RQ1 is addressed i.e. how different sets of event data and event logs affect process mining results. Then, the role of process mining in building simulation models is assessed by considering aggregated building blocks (entities, activities, control-flow, and resources) given the results of the case studies, presenting the answer to RQ2.

Two different datasets have been used for process mining and imported into Disco process mining. The software enables automatic discovery of process models based on imported data using optimised high-speed process discovery algorithm. The tool is available from https://fluxicon.com/disco/. Resulting process maps are intuitive and easy to work with, and the software enables dynamic process overview (lapse of time can also be animated using the model). The tool provides extensive reporting features, and a great number of process maps can be generated, however, due to page limit, only several are selected to be presented in the analysis. A downside is that the rendered maps cannot be easily customised, consequently being hard to read.

Background: process mining in smart mobility and higher education

Smart mobility

Popular use case in the smart city domain is the implementation of smart parking solutions, which contributes to the optimisation of peoples' time, reduction of fuel consumption and carbon dioxide emissions (Barriga et al., 2019). The architecture of smart parking solutions is frequently characterised by three components - sensors, communication protocols, and software solutions (ibid). As a part of the literature review, the keyword search has been conducted in the Scopus database to review the state-of-art of process mining in smart mobility, but no results have been found filtering the title option. An additional check has been done to explore whether any other type of mining has been discussed in the smart mobility context (smart + mobility + mining). Several studies matched the criteria of keyword search; however, none of them in reference to the process mining, but mostly focusing on data mining and data analysis using different algorithms. A relevant study by Anchal and Mittal (2019) surveyed the existing smart parking system by looking at how the received data from the parking sensors can be processed through filtering and data fusion. Also, they reviewed different data mining techniques that can help identify patterns in the IoT-enabled smart parking system. Aggarwal and Toshniwal (2018) studied the applicability and corresponding data mining techniques for smart mobility issues. With an objective to reduce the number of circulating vehicles, an example is provided by Lira (2019) who improves ride-matching algorithms with alternative destinations and identifies potential users for transportation services by exploring the content of posts on microblogging platforms. On a similar note, Zhong et al. (2016) compared urban mobility patterns in London, Singapore and Beijing using smart card data for the period of one-week focusing on understanding the regularities in patterns of transit use, but also found regularity in variability. Pronello et al. (2018) also analysed mobility patterns using smart card data mining, and in this way, improving the design of transport supply and mobility services to meet the user needs. As stated, no relevant studies have been found in reference to the process mining – this supporting the importance of the topic and its pioneer position in the smart city context.

Higher education

As event data are becoming more accessible and process mining techniques mature, an increasing number of researchers are looking into opportunities for a better understanding of educational processes. Educational Process Mining (EPM) is used in a wide range of settings, some of which are presented here. Devi and Suryakala (2014) present valuable cases of process mining in the context of higher education systems such as discovering extreme values in student outcomes, predicting student performance, identifying potential dropouts, and students in need of assistance. On a similar note, Mukala et al. (2015) use process mining techniques to monitor and analyse student learning habits based on data collected from Massive Open Online Courses (MOOCs) establishing that successful students always watch videos in the recommended sequence, while the opposite is true for unsuccessful students. Related to MOOCs, Umer et al. (2017) propose a process mining approach to enhance the student learning experience by combining different machine learning techniques with process mining features to measure the effectiveness of different techniques. The ultimate aim of that particular data-based approach is to help improve the student learning experience and reduce dropout rates in MOOCs. Another study focused on the visualisation of students' behaviour using process mining techniques to inspect student graduation probability by discovering, monitoring, and improving processes based on the event log and trace data (Douzali & Darabi, 2016). Van Der Aalst et al. (2015) even proposed that process cubes should be used as a way of organising event log data in a multidimensional data structure adapted for process mining to allow comparison of different process variants or different groups of cases.

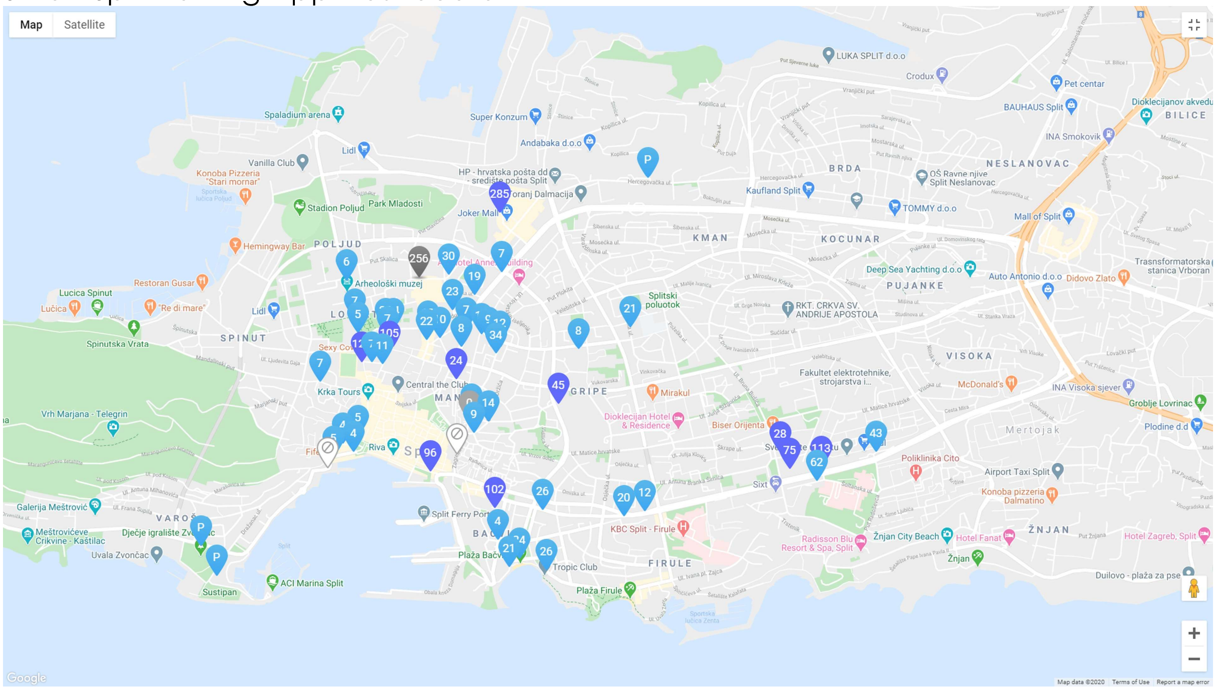
Setting the context of the two case studies

Smart mobility

In the city of Split in Croatia, by installing parking sensors and introducing the Smart Parking App, the objective was to contribute to reducing congestion, lowering exhaust levels, and reducing stress levels among drivers by making it easier for drivers to find parking spaces. The sensors detect a vacant or occupied parking space, and the status of the parking spaces is available at http://smart.splitparking.hr. The overview is presented in the form of pins on the map containing information about the number of vacant spaces (Figure 2) with the possibility to navigate the user to the nearest available or selected parking space (Split parking, 2019). Considering that sensors continually detect whether a place is vacant (0) or occupied (1) a considerable amount of data is recorded; however, with only several characteristics suitable for process mining. Each record contains the following data: id, id_parking_space, occupied, date, and vendor_id. Attributes "id" and "vendor_id" are unique and generated each time a parking space changes the status from occupied to vacant and vice versa. Attribute "occupied" is therefore binary. "Date" is a timestamp (formatted for example "2019-12-18 12:47:19") while "id_parking_space" represents a specific parking space in the city.

The issues of incomplete and raw data have been addressed already in the paper in referring to Van Der Aalst's (2015) guidelines for logging. Even though in this particular case, the data is not recorded in line with the guidelines, the case can serve as a demonstration that event data containing several attributes can contribute to the more straightforward conceptualisation and building of simulation models. There is a noticeable lack of information related to tracking a specific "id" since all "ids" are unique. Furthermore, there is only one activity that is the same for all "ids" – parking, while data about the resources (parking spaces) is only available as a binary value. The aim of the simulation can be related to optimisation scenarios, determination of peak load, or for example to identify poorly used parking spaces and release these to the general public (for example, for long term leases or similar), and in that context, the potential of process mining as a basis for building a simulation model is tested. Data processing was performed based on 114,931 unique records collected from a randomly selected parking lot in the city of Split within one year – 2019. As a part of the Extract, Transform, and Load (ETL) process, the necessary amendments were made to enrich the dataset with additional information related to time dimension so that it would contain vacancy/occupancy per weekdays and hours of the day.

Figure 2
Smart Split Parking App Dashboard



Source: Split parking, 2019

Higher education
Learning Management Systems (LMS) can be used to support traditional courses (to a lesser extent, usually for online publication of learning material), in a hybrid-learning and fully-online learning settings. The latter approach should be supported by a good number of features offered by the system, and this is the reason why, for this case, data from a fully-online course was used. Data is generated and recorded in Moodle LMS, a system used from the academic year 2008/2009 onwards at the Faculty of Economics in Split. As a part of the first-year course "Information Technology", students can access the e-course "Information Security" for eight weeks with the aim to raise their awareness of information security concepts and information security measures. Students can access course resources and activities (reading texts, watching videos, completing surveys, and more) in no particular order, in the sequence and dynamics that suit them best. To complete the course, they are required to achieve 70% of points in the test, without restrictions related to the maximum number of times to access the test or the delay between test attempts. For the study, data collected in the academic year 2018/19 is used from a Moodle report that records all the e-course activity data for all available resources. Selected data were exported to a file in .xls format suitable for further analysis in the process mining tool.

Analysing the case studies
Since the general aim was to evaluate the role of process mining results as inputs for discrete-event simulation modelling for both case study examples, here the simulation

model building blocks (entities, activities, control-flow, and resources) with regards to the potential of using process mining results are considered for each case.

Analysis of smart mobility case study

Entity. As mentioned above, in this particular case, id and vendor_id are unique and are generated each time a parking space is occupied or vacated; therefore, statistical processing of the entire set or a sample of data is possible. Monitoring an entity that can perform various activities in the system by occupying and releasing resources in this example is reduced to analysing a unique entity (car) for which data was not collected over time (longitudinally), and analysing only one activity (parking), as well as selecting a vacant parking space (resource). Given that arrivals are random, it is possible to select data where occupied = 0 and calculate an average time between the records that would correspond to the average time between arrivals of the car. For the time between the arrivals, it is possible to determine the mean of the exponential distribution. Actual entity arrival is a piece of critical information; however, in this example, data is collected for each entity arrival and departure, but under a unique id, i.e. arrival/departure sequence is not linked to the same entity. Concerning entities, it can be stated with certainty that only one entity will enter the system/parking space at a given time with each arrival. Descriptive statistics and statistical distributions for parking frequency per hour can be calculated using, for example, Input Analyzer tool by Rockwell Software: e.g. for a specific parking spot (labelled as 1062) Frequency = 10,519, Min = 1, Max = 23, Sample Mean = 12.1, Sample Std Dev = 6.92, Distribution is Beta, expression: 0.5 + 23 * BETA (0.888, 0.874), Square Error: 0.000982, Chi-Square Test (number of intervals = 23, degrees of freedom = 20, test statistic = 219, with corresponding p-value <0.005).

Activity. By looking at the parking activity whereby an entity occupies a parking space (1) for a certain period and then releases it without any particular expenditure of resources, it is possible to determine the duration of the activity, as well as certain specificities related to a particular parking space (resource). Figure 3 presents the frequency, relative frequency, median duration, and mean duration for the activity (parking) by hours of the day. It is interesting to compare the median and mean values of duration. Values of medians, ranging from 23 seconds to 4 minutes of occupancy and vacancy by hours of the day, indicate that 50% of data (frequency) less than or equal to this value, do not fall into the charged category (since the parking is free up to 5 minutes). Parking in the afternoon usually lasts longer than in the morning, taking into account the mean duration. Based on this data, it is possible to determine the type of distribution for the delay.

Control-Flow. Control-flow definition and gateway routing logic in this particular case is not satisfactorily supported. Although analyzing 15 parking spaces in this particular case, process map (Figure 4) shows the arrival of unique (anonymous) cars and the occupancy of 3 available parking spaces (labelled as 1053, 1062, and 1056). For example, parking space marked as 1062 has been occupied in 11,024 occasions by cars during the whole year, in contrast to 10,712 occasions when it was free. Considering that in this case, there is no sequential relationship between activities, analysing traces in the log process map does not contribute to the process of building a simulation model. However, by analysing the frequency information displayed on the map, it is possible to determine the routing logic, i.e. the probability of occupying a specific parking space (per day of the week or hour of the day).

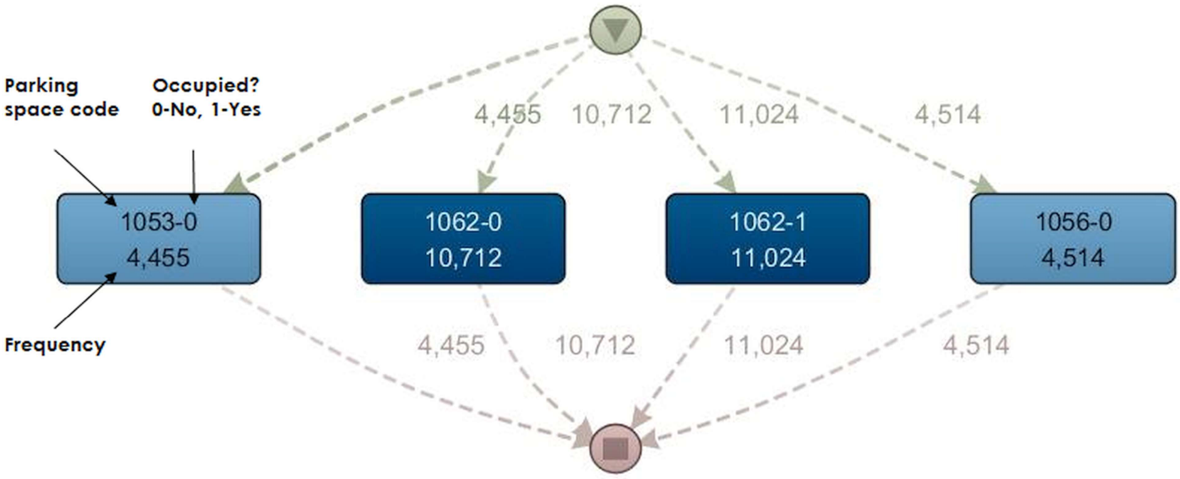
Figure 3
Frequency, Relative Frequency, Median Duration and Mean Duration for Parking by Hours of the Day

Hours of the day	Frequency	Relative frequency	Median duration	Mean duration
0	2,604	2.27%	28 secs, 136 millis	1 hour, 7 min
1	2,088	1.82%	28 secs, 43 millis	45 mins, 10 secs
2	1,595	1.39%	38 secs, 285 millis	24 mins, 59 secs
3	1,483	1.29%	42 secs, 41 millis	15 mins, 8 secs
4	2,263	1.97%	35 secs, 56 millis	21 mins, 37 secs
5	1,581	1.38%	41 secs, 283 millis	20 mins, 43 secs
6	3,564	3.10%	1 min	1 hour, 34 mins
7	4,379	3.81%	2 mins, 56 secs	1 hour, 31 mins
8	4,554	3.96%	4 mins, 1 sec	35 mins, 59 secs
9	5,173	4.50%	3 mins, 26 secs	31 mins, 36 secs
10	6,115	5.32%	2 mins, 46 secs	31 mins, 20 secs
11	6,780	5.90%	2 mins, 58 secs	35 mins, 37 secs
12	6,556	5.70%	2 mins, 52 secs	40 mins, 30 secs
13	6,129	5.33%	2 mins, 54 secs	44 mins, 58 secs
14	7,322	6.37%	3 mins, 1 sec	48 mins, 38 secs
15	7,314	6.36%	2 mins, 35 secs	56 mins, 43 secs
16	5,970	5.19%	1 min, 25 secs	1 hour, 3 mins
17	6,343	5.52%	1 min, 10 secs	1 hour, 46 secs
18	6,203	5.40%	1 min, 32 secs	1 hour, 52 secs
19	6,974	6.07%	1 min, 7 secs	1 hour, 9 mins
20	6,232	5.42%	47 secs, 194 millis	1 hour, 21 mins
21	5,180	4.51%	30 secs, 281 millis	1 hour, 21 mins
22	4,856	4.23%	23 secs, 71 millis	1 hour , 24 mins
23	3,673	3.20%	23 secs, 56 millis	1 hour, 25 mins

Source: Activity statistics from Disco process mining tool

Resources. In this case, there are 15 parking spaces representing resources that can perform the same activity (provide a timed parking service), with entities (cars) that seize-delay-release the resources. Specifically, resources are first allocated, followed by a process-delay, and then the allocated resource is released, which in this example does not affect resource expenditure but the generation of revenue. Collected data indicate that there are specific differences between the 15 available resources (Figure 5), which may serve to support the resource optimisation procedures. Namely, in the situation of resources underuse, or overcapacity even at peak loading times, there is potential for planning the termination of certain resources. In this case, it would lead to more efficient use of the public property, if for example, the management was to introduce green islands in these spaces.

Figure 4
Process Map Showing Occupancy of Selected Parking Spaces



*Source: Process m*ap created in Disco process mining tool

Figure 5
Frequency, Relative Frequency, Median Duration and Mean Duration for Activity Parking per Available Resources

Resource	Frequency	Relative frequency	Median duration	Mean duration
1062	21,736	18.91%	35 secs, 244 millis	6 mins, 44 secs
1056	8,464	7.36%	5 mins, 59 secs	56 mins, 19 secs
1053	8,074	7.03%	3 mins, 50 secs	59 mins, 8 secs
1064	7,773	6.76%	4 mins, 51 secs	1 hour, 59 secs
1065	7,466	6.50%	4 mins, 33 secs	1 hour, 1 min
1055	7,211	6.27%	3 mins, 17 secs	1 hour, 7 mins
1060	7,157	6.23%	1 min, 40 secs	59 mins, 8 secs
1054	6,885	5.99%	3 mins, 25 secs	1 hour, 9 mins
1057	6,721	5.86%	4 mins, 33 secs	1 hour, 11 mins
1052	6,550	5.70%	3 mins, 24 secs	1 hour, 13 mins
1051	6,552	5.67%	3 mins, 46 secs	1 hour, 8 mins
1058	5,532	4.81%	1 min, 19 secs	1 hour, 20 mins
1059	5,106	4.44%	1 min, 34 secs	1 hour, 32 mins
1061	4,997	4.35%	2 mins, 19 secs	1 hour, 18 mins
1063	4,727	4.11%	2 mins, 50 secs	1 hour, 16 mins

Source: Resource statistics from Disco process mining tool

Analysis of higher education case study

Entity. Process mining can be useful for analysing data about a series of different types of activities such as browsing, deleting, modifying, adding content within an e-course. These activities are generated in the learning process by the frequent involvement of students throughout the semester and their interaction with the learning material. In this second case, the process begins with student enrolment in the course and continues with several different activities denoting students' progress in the e-course. All the activities performed by the entity "student" are automatically recorded to the system, so it is possible to monitor their progress while attending the e-course. The time component is recorded as accurate entity arrival in the format "22/03/18, 13:28" and is one of the critical factors in measuring process performance. Descriptive statistics and distribution for the frequency of access to the e-course and the use of specific resources within the e-course are analysed using the Input Analyzer tool by Rockwell Software, and a part of it is presented in Table 3. It is noteworthy that the distribution given here, considering it is based on a smaller data segment and that there are large variations in the observed data, can only serve to indicate the potential for the use of this particular data in simulation modelling.

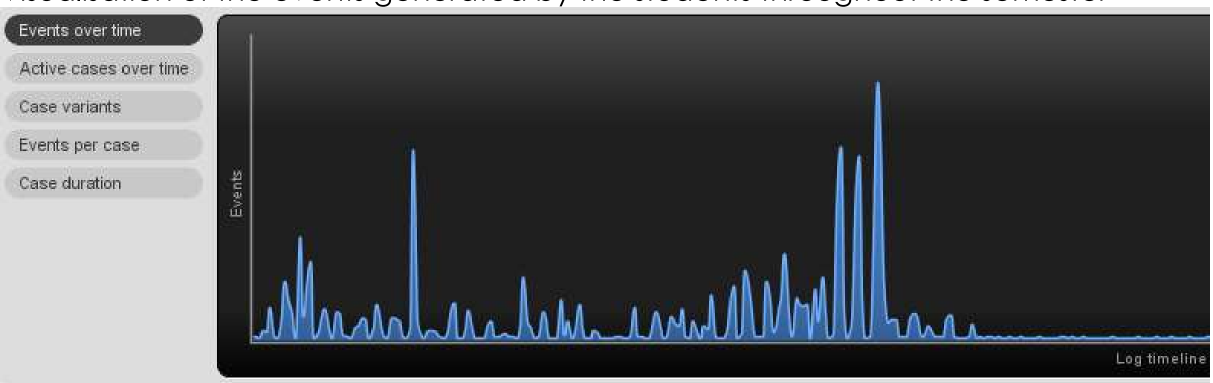
Table 3
Input Data Analysis of the LMS e-Course Access Data

Resource	Count	Min	Max	Mean	StD	Distribution	Expression
Choice	230	1	15	6,52	2,51	Normal	NORM(6.52, 2.51)
File	222	1	11	3,64	1,61	Gamma	0.5 + GAMM(0.745, 4.21)

Source: Moodle Event Log

Activity. The entity "student" performs educational activities (reading, writing) by interacting with certain resources in the e-course for a fixed time and then releasing them and moving on to the next activity or leaving the e-course. By using process mining tools, it is possible to determine the duration of the activities. Process discovery builds process models drawing from the data recorded in the event log. An event log can be used as an input to numerous process mining algorithms to visualise and uncover the actual behaviour (sequential steps) of students. Disco provides a complete set of process metrics for activities and paths absolute frequency: case frequency, the maximum number of repetitions, total duration, mean duration, maximum duration. The models illustrated here show the activities and their performance in the form of the absolute frequency of individual activities and the transitions between them. The resulting events over time (Figure 6) can also be tracked and used for validation of a simulation model. The tool calculates the total number of events (33,149), number of specific cases (366), number of specific activities (73), median case duration (42.2 hours) and mean case duration (12.8 days).

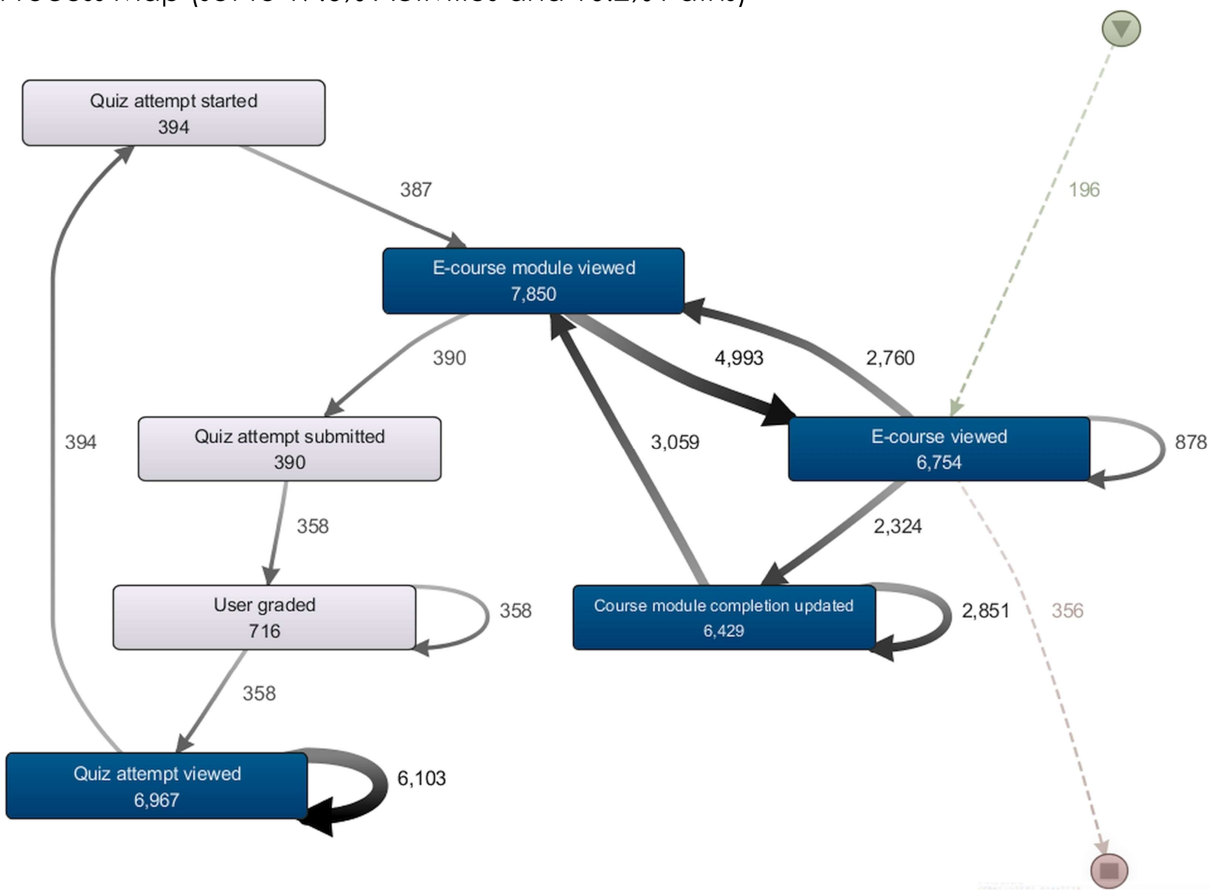
Figure 6
Visualisation of the events generated by the students throughout the semester



Source: Overview diagram "Events Over Time "created in Disco process mining tool

Control-Flow. Control-flow definition and gateway routing logic are well-supported in this case. The process map illustrates students accessing the e-course, performing different activities using different resources, and leaving the e-course upon completion. The Disco tool allows fine-tuning of the percentage of displayed activities and paths. Process map that represents 100% of activities and 56.7% of paths is not informational due to its complexity, and it does not serve as a useful input for the simulation modelling process itself. However, reducing the activity view to, for example, 17.6% and the path to 10.2% (Figure 7), provides a perspective that can support control-flow and routing logic modelling. In this particular case, there are several sequential relationships between activities; therefore, the analysis of the traces in the log files and the process map can contribute to the process of building a conceptual and computer simulation model. Also, by analysing the frequency information displayed on the map, it is possible to determine the routing logic, that is, the probability of an occurrence of a specific activity and the routing of entities in selecting a subsequent activity.

Figure 7
Process Map (set to 17.6% Activities and 10.2% Paths)



Source: Process map created in Disco process mining tool

Resources. In this scenario, the frequencies and relative frequencies for the resources are: System (f = 16.535; 49.88%), Test (f = 9.307, 28.08%), Page (f = 3.871, 11,68%), Choice (f = 2,073, 6.25%), File (f = 807, 2.43%), Dictionary (f = 521, 1.57%), and so on, as is presented in Figure 8. Statistics are also automatically calculated for the use of all resources (11), such as minimum frequency (1), median frequency (521), mean frequency (3,013.55), maximum frequency (16,535), and frequency standard deviation (5,294.67).

Figure 8
Resources affected by the students in Moodle system

Resource	Frequency	Relative frequency
System	16,535	49.88%
Test	9,307	28.08%
Page	3,871	11.68%
Choice	2,073	6.25%
File	807	2.43%
Dictionary	521	1.57%
Student report	18	0.05%
Forum	10	0.03%
Recycle bin	3	0.01%
Report	3	0.01%

Source: Statistics from Disco process mining tool

Discussion

The paper supports the findings that there are still significant research and practical challenges in collecting event data for process mining, and consequently, in integrating the two fields – process mining and discrete-event simulation. The potential for the use of process mining in building a simulation model was elaborated and reviewed based on a related study by Martin et al. (2015) per each of the aggregated simulation building blocks: entities, activities, control-flow, and resources. The concept brought forth in Table 2 was then developed further and empirically tested, and the results were broken down per simulation building block in two noticeably different case studies. It was made apparent that the success of process mining depends on the availability of event logs that are clearly defined and need to refer to a case (i.e. process instance) and an activity (i.e. step in the process), this way answering RQ2. Not all (or better even: not a lot of) information systems can simply generate event logs that match presented 15 guidelines and recommendations, while most generate raw data that needs to be linked to cases and activities when creating or using process models (Van Der Aalst, 2015).

The two case studies confirmed this particular situation. Process mining results based on the parking data represent a poor-data case, while the process mining on LMS data can be considered a rich-data case. As investigated within RQ1, the results of PM analysis point to the conclusion that the quality of event datasets affects PM results. In the first case, with limited data, the findings support the conclusions noted by the IEEE Task Force on Process Mining (2011), where event data incompleteness has been noted as one of the important challenges within "Finding, Merging, and Cleaning Event Data" challenge in The Process Mining Manifesto. R'bighui and Cho (2017) investigated whether the defined challenges have been resolved in the following years, and found that this area is still interesting to the research community, however, due to a variety of different sources of event data, there is still a need for additional research. For example, in the pursuit to rediscover the process models from small incomplete event logs, Leemans et al. (2014) successfully introduced a new algorithm based on the inductive miner and probabilistic behavioural relations, that are less sensitive to incomplete logs. As the quality of process mining results directly depends on event logs, finding, merging and cleaning event logs were reported as the first challenge in PM manifesto understandably so.

On the contrary, rich-data cases offer more possibilities and promise in terms of delivering more conclusive and relevant PM results. As such they represent higher-quality input for simulation modelling, as evidenced in studies, for example, by Phan et al. (2019), Tamburis (2019), Jadrić (2019), Abohamad et al. (2017), Zhou et al. (2014), and Rozinat et al. (2009), where valuable information has been collected, i.e. potential for improvements in the processes were identified. In the study by Nakatumba et al. (2012), it has been confirmed that combining process mining with simulation leads to better understanding, modelling, and improving real-life business processes.

Apart from the contribution in the area of simulation modelling (by providing a framework for more objective building of simulation models), this paper provides contributions in two other domains considering the two case studies:

Smart City (SC) domain is characterised as an area of open and user-driven innovation (Schaffers et al., 2011) thus being a good platform for experimentation and implementation of new technologies, products, and services. Smart mobility as one of the SC pillars aims to increase the efficiency of urban transportation (Giffinger & Gudrun, 2010) and to reduce traffic congestion and harmful environmental influences, in this way positively influencing the quality of citizens' lives. The practical implications of designing a simulation study based on the results of process mining in the parking domain are numerous. For example, business value can be found in the analysis of what-if scenarios of using a mobile application such as the one presented in the paper, e.g. by introducing pay-per-use and not per fixed (pre-paid) period, or by deliberating code-sharing in case a user has not used up all the pre-paid time promoting in that way concepts of the smart and sharing economy. In this specific case of smart parking, the service is handled by two companies, the first one being a city-owned company that is also the owner of the infrastructural assets (including sensors), and the other one handling ticketing-payment process (outsourced activity). Information on ticketing-payment was not available to the authors, and therefore cross-organizational mining has not been performed, with a view to do so in the future.

Educational Process Mining (EPM) is a relatively new research area within Educational Data Mining (EDM) that aims to uncover valuable patterns by using logs collected from educational settings to analyse and provide a visual representation of educational processes and to provide better insights (Bogarín et al., 2018). Results-based process mining of educational data, such as the one presented in this case, followed the general concept of process mining by providing the basis for discovering, monitoring, and improving real processes by extracting new knowledge from logs that are automatically recorded by frequently used information systems. The practical implications of conceptualising a simulation study based on the results of a process mining case in an educational environment are numerous. Based on the process-related knowledge discovery from extensive records of educational events in the form of new process models that track key performance indicators, it is possible to: analyse the “real” educational processes and their alignment with the curricula; improve the performance indicators of educational processes such as execution time, bottlenecks, decision points, and similar; personalise educational processes by recommending the most suitable learning units or learning paths to students (all depending on their profiles, their preferences or target skills) (Grigorova et al., 2017). In addition to this, by building and using a simulation model based on LMS data, the testing of different what-if scenarios is possible without interfering with the actual educational process.

Conceptual models, being an important part of business process simulations, are usually based on information sources such as process documentation, expert interviews, and observations. There are many issues with these traditional sources, as they differ in formats and often include the problem of biases, which may lead to the discrepancy between the simulated model and the reality (Rozinat et al., 2009). As the behaviour of business processes is increasingly registered in the event logs of most of the information systems today (customer relationship management systems, enterprise resource planning systems, and other), there are growing opportunities to introduce event log knowledge, extracted from log files using process mining techniques, as an additional input to building simulation models. The rationale behind the new research on the links between process mining and simulation modelling is clear: event logs are files where process-aware information systems are registering the information about the actual behaviour of the process (as successfully illustrated by Martin et al., 2014). In order to extract useful knowledge from an event log, process mining techniques have to be employed. The results, in the form of new knowledge and insights, can then be validated through traditional information sources, but also by staff members and experts as they may locate errors and inaccuracies in data sets, which are the basis for process mining (Abohamad et al., 2017).

Conclusion

This paper provided an overview of the possible scenarios in which process mining concepts could be used to support simulation model construction. The link has been demonstrated per each of the aggregated simulation building blocks: entities, activities, control-flow, and resources. The whole process is presented in the paper for two different models. The results related to activities, control-flow and resources for the simulation entities "car" and "student", together with the timestamps, descriptive statistics, and distributions can be used in the specific phases (most of all in the input phase) of the conceptual and discrete-event simulation model development, as well as in the validation procedures.

In addition to this, the paper emphasised the importance of complete event log data, then suitable for process mining in different domains. Process maps and statistical calculations in the cases of smart parking and LMS data were presented using the Disco tool, and along with the implications of the results for each of the building blocks.

In addressing one of the key challenges that motivated the research – combining process mining with simulation (RQ2), this study addressed other challenges noted in the PM manifesto (IEEE Task Force, 2011; Van Der Aalst et al. 2012) as well: from the category of finding, merging and cleaning event data – incomplete event data (RQ1). Building on the existing studies (mostly to support the framework for mapping process mining with simulation provided here, e.g. Martin et al. (2015) and Abohamad et al. (2017), this study contributed both to process mining and discrete-event simulation research.

In the smart city context, smart mobility, in particular, presents an important area with the potential for the use of discrete-event simulation (Jadrić et al., 2019). Coupled with process mining the research area offers a suitable innovation environment. Practical implications in the smart parking case refer to possible improvements in business processes by implementing better use of (public) resources, but also point to the constraints of information system in use, in terms of the non-existence of specific event data. In the case of higher education, PM use on LMS data can contribute to a better prediction of student behaviour, new modalities of work balanced with students' needs and preferences. The smart context is highlighted in this domain as well; according to Waheed et al. (2018), there is a need to integrate learning analytics research with multidisciplinary smart education and smart library service in the future.

Limitations of the study include sub-optimal data used in the two case studies, and further research is required to offer and substantiate improvements in future strategies and action plans. In addition, cross-organizational process mining techniques have not been addressed in the case of smart parking data to get the complete visualisation of the process, which would certainly be the aim of the future case study. To summarize, future research on combining process mining and simulation needs to be extended further to provide a stronger theoretical and empirical background of PM, which will then be efficient in constructing the structure of the simulation model, equal to real-life.

References

1. Abohamad, W., Ramy, A., Amr, A. (2017), "A hybrid process-mining approach for simulation modelling", 2017 Winter Simulation Conference, 3-6 December, IEEE, Las Vegas, pp. 1527-1538.
2. Aggarwal A., Toshniwal D. (2018), "Data mining techniques for smart mobility—a survey", in Sa, P., Bakshi, S., Hatzilygeroudis, I., Sahoo, M. (Eds.), Advances in Intelligent Systems and Computing, Vol. 709, Springer, Singapore, pp. 239-249.
3. Agrawal, R., Gunopulos, D., Leymann, F. (1998), "Mining process models from workflow logs", in Schek, H. J. (Ed.), 6th International Conference on Extending Database Technology: Advances in Database Technology, Vol. 1377, Springer, Berlin, Heidelberg, pp. 467-483.
4. Anchal, Mittal, P. (2019), "Data mining techniques for IoT enabled smart parking environment: survey", International Journal of Advanced Trends in Computer Science and Engineering, Vol. 8, No. 4, pp. 1688-1697.
5. Baier, T., Mendling, J. (2013), "Bridging abstraction layers in process mining by automated matching of events and activities", in Daniel, F., Wang, J., Weber, B. (Eds.), Business Process Management. Lecture Notes in Computer Science, Vol. 8094, Springer, Berlin, Heidelberg, pp. 17-32.
6. Barriga, J. J., Sulca, J., León, J. J., Ulloa, A., Portero, D., Andrade, R., Yoo, S. G. (2019), "Smart parking: a literature review from the technological perspective", Applied Sciences, Vol. 9, No. 21, 4569.
7. Bogarín, A., Cerezo, R., Romero, C. (2018), "A survey on educational process mining", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol. 8, No. 1, e1230.
8. Cattafi, M., Lamma, E., Riguzzi, F., Storari, S. (2010), "Incremental declarative process mining", in Szczerbicki E., Nguyen N. T. (Eds.) Smart Information and Knowledge Management, Vol. 260, Springer, Berlin, pp. 103-127.
9. Cook, J. E., Wolf, A. L. (1998), "Discovering models of software processes from event-based data", ACM Transactions on Software Engineering and Methodology, Vol. 7, No. 3, pp. 215-249.
10. de Medeiros, A., Weijters, A., Van Der Aalst, W. (2005), "Genetic process mining: a basic approach and its challenges", in Bussler, C. J., Haller, A. (Eds.), Business Process Management Workshops. Lecture Notes in Computer Science, Vol 3812, Springer, Berlin, Heidelberg, pp. 203-215.
11. Devi, K. L., Suryakala, M. (2014), "Educational process mining-different perspectives", IOSR Journal of Computer Engineering, Vol. 16, pp. 57-60.
12. Douzali, E., Darabi, H. (2016), "A case study for the application of data and process mining in intervention program assessment", 123rd ASEE Annual Conference & Exposition, 26-29 June, American Society for Engineering Education, New Orleans, 15895.
13. Dustdar, S., Hoffmann, T., Van Der Aalst, W. (2005), "Mining of ad-hoc business processes with TeamLog", Data & Knowledge Engineering, Vol. 55, No. 2, pp. 129-158.
14. Giffinger, R., Gudrun, H. (2010), "Smart cities ranking: an effective instrument for the positioning of the cities", ACE: Architecture, City and Environment, Vol. 4, No. 12, pp. 7-26.
15. Grigorova, K., Malysheva, E., Bobrovskiy, S. (2017), "Application of data mining and process mining approaches for improving e-learning processes", 3rd International Conference on Information Technology and Nanotechnology, 25-27 April, ITNT, Samra, Vol. 1903, pp. 115-121.

16. IEEE Task Force on Process Mining. (2011), Process mining manifesto, available at: https://www.win.tue.nl/ieeetfpm/downloads/Process%20Mining%20Manifesto.pdf (22 April 2020)
17. Jadrić, M. (2019), "Framework for discrete-event simulation modelling supported by LMS data and process mining", in Zadnik Stirn, L., Kljajić Borštnar, M., Žerovnik, J., Drobne, S., Povh, J. (Eds.), 15th International Symposium on Operational Research, 25-27 September, Slovenian Society Informatika, Bled, pp. 225-230.
18. Jadrić, M., Ćukušić, M., Pavlić, D. (2019), "Review of discrete simulation modelling use in the context of smart cities", in SSRCI - Smart, Sustainable and Resilient Cities and Infrastructures, 20-24 May, Opatija, Croatia.
19. Leemans, S. J. J., Fahland, D., Aalst, W. M. P. (2014), "Discovering block-structured process models from incomplete event logs", in Ciardo G., Kindler E. (Eds.), Application and Theory of Petri Nets and Concurrency. Lecture Notes in Computer Science, Springer, Tunis, Vol. 8489, pp. 91-110.
20. Lira, V. C. M. (2019), Mining human mobility data and social media for smart ride sharing, Doctoral dissertation, Universidade Federal de Pernambuco, available at: https://repositorio.ufpe.br/bitstream/123456789/34515/1/TESE%20Vinicius%20Cezar%20Monteiro%20de%20Lira.pdf (3 February 2020)
21. Liu, S. (2015), Integrating process mining with discrete-event simulation modeling, Master thesis, Brigham Young University, available at: https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=6734&context=etd (3 February 2020)
22. Martin, N., Depaire, B., Caris A. (2014), "Event log knowledge as a complementary simulation model construction input", 4th International Conference on Simulation and Modeling Methodologies, Technologies and Applications, 28-30 April, INSTICC, Vienna, pp. 456-462.
23. Martin, N., Depaire, B., Caris A. (2015), "The use of process mining in business process simulation model construction", Business & Information Systems Engineering, Vol. 58, No. 1, pp. 73-87.
24. Mukala, P., Buijs, J., Leemans, M., Van Der Aalst, W. (2015), "Learning analytics on coursera event data: A process mining approach", in Ceravolo, P., Rinderle-Ma, S. (Eds.), 5th International Symposium on Data-driven Process Discovery and Analysis, 9-11 December, CEUR, Vienna, Vol. 1527, pp. 18-32.
25. Nakatumba, J., Westergaard, M., Aalst, W. M. P. (2012), "Generating event logs with workload-dependent speeds from simulation models", in Bajec, M., Eder, J. (Eds.), Advanced Information Systems Engineering Workshops. Lecture Notes in Business Information Processing, 25-26 June, Springer, Berlin, Heidelberg, Gdańsk, Vol. 112, pp. 383-397.
26. Phan, R., Augusto, V., Martin, D., Sarazin, M. (2019), "Clinical pathway analysis using process mining and discrete-event simulation: an application to incisional hernia", 2019 Winter Simulation Conference, 8-11 December, IEEE, National Harbor, pp. 1172-1183.
27. Pospišil, M., Hruška, T. (2012), "Business process simulation for predictions with focus on decision mining and execution time of tasks", 2nd International Conference on Business Intelligence and Technology, 22-27 July 2011, IARIA, Valencia, pp. 14-18.
28. Pronello, C., Longhi D., Gaborieau J. B. (2018), "Smart card data mining to analyze mobility patterns in suburban areas", Sustainability, Vol. 10, No. 10, 3489.
29. R'bigui, H., Cho, C. (2017), "The state-of-the-art of business process mining challenges", International Journal of Business Process Integration and Management, Vol. 8, No. 4, pp. 285-303.
30. Rojas, E., Munoz-Gama, J., Sepúlveda, M., Capurro, D. (2016), "Process mining in healthcare: a literature review", Journal of Biomedical Informatics, Vol. 61, pp. 224-236.
31. Rovani, M., Maggi, F. M., De Leoni, M., Van Der Aalst, W. M. P. (2015), "Declarative process mining in healthcare", Expert Systems with Applications, Vol. 42, No. 23, pp. 9236-9251.
32. Rozinat, A., Mans, R. S., Song, M., Van Der Aalst, W. M. P. (2009), "Discovering simulation models", Information Systems, Vol. 34, No. 3, pp. 305-327.

33. Schaffers, H., Komninos, N., Pallot, M., Trousse, B., Nilsson, M., Oliveira, A. (2011), "smart cities and the future internet: towards cooperation frameworks for open innovation", in Domingue, J., Galis, A., Gavras, A., Zahariadis, T., Lambert, D., Cleary, F., Daras, P., Krco, S., Müller, H., Li, M.-S., Schaffers, H., Lotz, V., Alvarez, F., Stiller, B., Karnouskos, S., Avessta, S., Nilsson, M. (Eds.), Future Internet Assembly 2011: Achievements and Technological Promises, Vol 6656. Springer, Berlin, Heidelberg, pp. 431-446.
34. Song, M., Günther, C., Van Der Aalst, W. (2009), "Trace clustering in process mining", in Ardagna, D., Mecella, M., Yang, J. (Eds.), Business Process Management Workshops. Lecture Notes in Business Information Processing, 1-4 September, Milano, Springer, Berlin, Heidelberg, Vol 17, pp. 109-120.
35. Split Parking. (2019), "Parking locations", available at: http://smart.splitparking.hr (23 April 2020)
36. Tamburis, O. (2019), "Bridging the gap between process mining and des modeling in the healthcare domain", 7th E-Health and Bioengineering Conference, 21-23 November, IEEE, Iasi, 8969912.
37. Tiwari, A., Turner, C. J., Majeed, B. (2008), "A review of business process mining: state of the art and future trends", Business Process Management Journal, Vol. 14, No. 1, pp. 5-22.
38. Turner, C. J., Tiwari, A., Olaiya, R., Xu, Y. (2012), "Business process mining: from theory to practice", Business Process Management Journal, Vol. 18, No. 3, pp. 493-512.
39. Umer, R., Susnjak, T., Mathrani, A., Suriadi, S. (2017), "On predicting academic performance with process mining in learning analytics", Journal of Research in Innovative Teaching & Learning, Vol. 10, No. 2, pp.160-176.
40. Van Der Aalst, W. M. P. (2011), Process Mining: Discovery, Conformance and Enhancement of Business Processes, Springer Science & Business Media, Heidelberg, Berlin.
41. Van Der Aalst, W. M. P. (2012), "Process mining: overview and opportunities", ACM Transactions on Management Information Systems", Vol. 3, No. 2, pp. 1-17.
42. Van Der Aalst, W. M. P. (2015), "Extracting event data from databases to unleash process mining", available at: http://wwwis.win.tue.nl/~wvdaalst/publications/p817.pdf (3 February 2020)
43. Van Der Aalst, W. M. P., Guo, S., Gorissen, P. (2015), "Comparative process mining in education: an approach based on process cubes", Lecture Notes in Business Information Processing, Vol. 203, pp. 110-134.
44. Van Der Aalst, W. M. P., Rubin, V., Verbeek, H. M. W., Van Dongen, B. F., Kindler, E., Gunther, C. W. (2010), "Process mining: a two step approach to balance between underfitting and overfitting", Software & Systems Modelling, Vol. 9, No. 1, pp. 87-111.
45. Van Der Aalst, W. M. P., Weijters, A. J. (2004), "Process mining: a research agenda", Computers in Industry, Vol. 53, No. 3, pp. 231-444.
46. Wagner, G., Seck, M., McKenzie, F. (2016), "Process modeling for simulation: observations and open issues," 2016 Winter Simulation Conference, 11-14 December, IEEE, Washington, DC, pp. 1072-1083.
47. Waheed, H., Hassan, S. U., Aljohani, N. R., Wasif, M. (2018), "A bibliometric perspective of learning analytics research landscape", Behavior and Information Technology, Vol. 37, No. 10-11, pp. 941-957.
48. Wang, Y. (2018), "An integrative process mining approach to mine discrete event simulation model from event data", Doctoral thesis, Université de Bordeaux, available at: https://tel.archives-ouvertes.fr/tel-01952795/document (3 February 2020)
49. Weijters, A., Van Der Aalst, W. (2001), "Process mining: discovering workflow models from event-based data", available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.7.7434 (3 February 2020)
50. Weijters, A., Van Der Aalst, W. (2003), "Rediscovering workflow models from event-based data using little thumb", Integrated Computer-Aided Engineering, Vol. 10, No. 2, pp. 151-162.
51. Weijters, A., Van Der Aalst, W., Medeiros, A. (2006), "Process mining with the heuristics miner algorithm", Technical Report No. 166, Technische Universiteit Eindhoven, Netherland. Available at:

https://pdfs.semanticscholar.org/1cc3/d62e27365b8d7ed6ce93b41c193d0559d086.pdf (20 April 2020)
52. Weske, M. (2012), Business Process Management: Concepts, Languages, Architectures, Springer, New York.
53. Zakarija, I., Škopljanac-Mačina, F., Blašković B. (2020), "Automated simulation and verification of process models discovered by process mining", Automatika, Vol. 61, No. 2, pp. 312-324.
54. Zhong, C., Batty, M., Manley, E., Wang, J., Wang, Z., Chen, F., Schmitt, G. (2016), "Variability in regularity: mining temporal mobility patterns in London, Singapore and Beijing using smart-card data", PLoS ONE, Vol. 11, No. 2. DOI: 10.1371/journal.pone.0149222
55. Zhou, Z., Wang, Y., Lin, L. (2014), "Process mining based modeling and analysis of workflows in clinical care – a case study in a Chicago Outpatient Clinic", 11th IEEE International Conference on Networking, Sensing and Control, 7-9 April, IEEE, Miami, pp. 590-595.

About the authors

Mario Jadrić, PhD is an Associate Professor at the Faculty of Economics, Business, and Tourism (FEBT), University of Split, Department of Business Informatics. He holds a PhD from the University of Zagreb, Faculty of Organization and Informatics. (Co)-author of more than 60 scientific papers on the topic of e-learning, technological innovations, business processes improvement and user perception of new technologies. He teaches e-learning in Business Environments, Business Process Simulation and Business Information Systems at undergraduate and graduate level and Computer Simulation at the PhD level. He is a member of the research team of MIS4SC project financed by Croatian Science Foundation and has participated in several international and national projects. The author can be contacted at **jadric@efst.hr**

Ivana Ninčević Pašalić is a Research and Teaching Assistant at FEBT's Department of Business Informatics, and a PhD candidate at the University of Split. She teaches lab classes for the courses Multidimensional Information Systems and Management of IT projects. Her previous working positions include being Internal Auditor for a US-based international organization and Operations Manager at a Croatian private company. She is also a member of the research team of MIS4SC project. Within the framework of the project, her research is focused on G2C and C2G interactions in smart cities, with a special focus on citizen engagement via different information and communication technologies. The author can be contacted at **nincevic@efst.hr**

Maja Ćukušić, PhD is an Associate Professor at FEBT, University of Split, Department of Business Informatics. Before joining academia, she worked for three years in a Dutch company in the design and implementation of complex ICT solutions for global markets. Currently, she is the manager and the principal investigator of the project financed by the Croatian Science Foundation on the topic of smart cities, leading the group of three PhD students and five international researchers. She is also a key expert for e-learning in the SEA-EU European University Alliance. In her teaching activities (courses OLAP systems, ERP systems, E-business, E-learning), she uses examples and studies on the use of business intelligence technologies, visualization tools (open data), smart devices and similar. The author can be contacted at **mcukusic@efst.hr**