# Event Log Knowledge as a Complementary Simulation Model Construction Input

Niels Martin[1], Benoît Depaire[1,2] and An Caris[1,2]
[1]*Hasselt University, Agoralaan – Building D, 3590 Diepenbeek, Belgium*
[2]*Research Foundation Flanders (FWO), Egmontstraat 5, 1000 Brussels, Belgium*

Keywords: Business Process Simulation, Event Log Knowledge, Process Mining, Simulation Model Construction Inputs, Conceptual Framework.

Abstract: Business process simulation models are typically built using model construction inputs such as documentation, interviews and observations. Due to issues with these information sources, efforts to further improve the realism of simulation models are valuable. Within this context, the present paper focuses on the use of process execution data in simulation model construction. More specifically, the behaviour of contemporary business processes is increasingly registered in event logs by process-aware information systems. Knowledge can be extracted from these log files using process mining techniques. This paper advocates the addition of event log knowledge as a model construction input, complementary to traditional information sources. A conceptual framework for simulation model construction is presented and the integration of event log knowledge during the modeling of particular simulation model building blocks is outlined. The use of event log knowledge is demonstrated in a simulation of the operations of a roadside assistance company.

## 1 INTRODUCTION

Business process simulation (BPS) refers to the imitation of business process behaviour through the use of a simulation model (Melão and Pidd, 2003). By mimicking the real system, simulation can contribute to the analysis and potentially the improvement of business processes (Rozinat et al., 2008).

Simulation models are typically created by simulation experts based on insights from traditional information sources such as process documentation, interviews and observations. Issues with these information sources, as will be outlined in section 2.1, may contribute to the discrepancy between the constructed simulation model and reality (Rozinat et al., 2009). Consequently, efforts to further improve the realism of simulation models are valuable as they will enhance the representativeness of analysis results and hence its relevance for management support. The present paper focuses on complementing traditional information sources with insights from data depicting actual process behaviour in an effort to construct more realistic simulation models.

With regards to process behaviour data, note that contemporary business processes are increasingly supported by process-aware information systems (PAIS) such as customer relationship management systems and enterprise resource planning systems. This type of systems register highly relevant information on the actual behaviour of the process under consideration in files called event logs. These files can be analyzed through the use of process mining techniques (van der Aalst, 2011). To date, this source of information is not used altogether or tends to be underexploited in BPS model construction.

This paper presents a conceptual framework for simulation model construction taking both traditional information sources and event log knowledge into account as complementary inputs. Moreover, some modeling aspects are detailed where event log analysis can deliver a contribution.

The remainder of this paper is structured as follows. In the second section, the aforementioned conceptual framework is justified and presented. The third section highlights some simulation model building blocks where event logs can provide valuable insights during the modeling process. An


Martin N., Depaire B. and Caris A..
Event Log Knowledge as a Complementary Simulation Model Construction Input.
DOI: 10.5220/0005100404560462
In *Proceedings of the 4th International Conference on Simulation and Modeling Methodologies, Technologies and Applications* (SIMULTECH-2014), pages 456-462
ISBN: 978-989-758-038-3

illustrative simulation study is outlined in the fourth section. The paper ends with a discussion and conclusions.

# 2 CONCEPTUAL FRAMEWORK FOR SIMULATION MODEL CONSTRUCTION

This section introduces a conceptual framework regarding BPS model construction. The model is justified in section 2.1 to support its relevance. Section 2.2 presents the actual framework.

## 2.1 Framework Justification

Various information sources are used to gain insights in a business process when constructing a simulation model. These typically include process documentation, interviews with business experts and the observation of real-life processes (Rozinat et al., 2009).

However, the obtained information from these sources might be biased. According to Mărușter and van Beest (2009), process documentation might deviate from real-life process behaviour. Interviews with business experts can result in contradictory information (Vincent, 1998) and their perception, as human perception in general, tends to be biased to a certain extent (Pronin, 2007). Observational data, in their turn, could suffer from the Hawthorne effect, which refers to the performance increase of staff members due to the attention they receive as their actions are observed (Brysbaert, 2006).

These issues contribute to a discrepancy between modelled and real process behaviour (Rozinat et al., 2009). Consequently, further efforts to improve the realism of simulation models are required. One approach could be the use of process execution data embedded in event logs. These data reflect the real-life behaviour of the process and can provide valuable insights for model construction purposes.

To date, the use of process execution data in BPS models is often limited to model parameterization (Law, 2007), e.g. the estimation of entity arrival rates and activity service times. The present paper advocates a broader and more systematic use of process execution data as a complementary input in simulation model construction. Besides distribution estimation, event data can also provide information about the order of activities, the model's decision logic, etc.

## 2.2 Conceptual Framework

This section outlines a conceptual framework towards simulation model construction which introduces event log knowledge as an additional input. A visual representation of the framework is presented in figure 1.

When disregarding event log knowledge, a simulation model is constructed based on information sources such as business documents, interviews with business experts and observations. These model construction inputs are aligned and analyzed to come up with a partial simulation model, i.e. a simulation model 'under construction'.

When knowledge hiatuses are identified during the construction, the modeler can return to the inputs and e.g. conduct additional expert interviews. This is visualized by connecting the partial simulation model to the model construction inputs in figure 1. Another issue that might lead to the use of this connection is the presence of non-tolerable differences between process performance metrics in the partial simulation model on the one hand and reality on the other hand. Deviating values of e.g. average throughput time will necessitate the addition or adjustment of particular simulation model parameters or other model components. Consequently, a return to the model construction inputs will be needed. The two outlined issues show that it might require several iterations before the simulation model is completed and the modeler can pursue towards running the final model and interpreting the results.

The conceptual framework introduces an additional model construction input: event log knowledge. To understand the origins of this extra input, note that contemporary business processes are increasingly supported by systems such as customer relationship management systems and enterprise resource planning systems. This type of systems are called process-aware information systems (PAIS) because the process notion is embedded in them. In contrast, a standard e-mail system can support a business process, but is not aware of the process it backs. Consequently, it is not a PAIS (van de Aalst, 2011). To be able to support a business process, the structure of the PAIS has to mimic the real-life process structure to a large extent.

From PAIS, event logs can be retrieved which capture process execution data. An illustrative example of an event log of a roadside assistance company is given in table 1. Each line in the log file corresponds to a single event, e.g. the registration of a customer request. Additional information on the event can be recorded such as a timestamp and the resource which is associated with the event. Events should be related to a particular case, e.g. a customer request. In simulation terms, a case corresponds to an entity flowing through the process.

Table 1: Illustration of an event log.

| *Case* | *Event* | *Timestamp* | *Activity* | *Resource* | *…* |
|---|---|---|---|---|---|
| 1 | 72 | 11/05/2014 12:03 | Request receipt | John | … |
| | 73 | 11/05/2014 12:05 | Request transmission | John | … |
| | 78 | 11/05/2014 12:28 | Request acceptance | Frank | … |
| | … | … | … | … | … |
| 2 | 135 | 11/05/2014 14:12 | Request receipt | Alice | … |
| | 138 | 11/05/2014 14:17 | Request transmission | Patrick | … |
| | 164 | 11/05/2014 14:41 | Patrolman departure | Peter | … |
| | … | … | … | … | … |

As the event log reflects the actual behaviour of the real-life process, it contains valuable intelligence for the construction of simulation models. In order to extract usable knowledge from an event log, process mining can be applied, which refers to the extraction of process knowledge from event logs.

This might relate to the order of activities in the process, the identification of resource roles, etc. (van de Aalst, 2011). Some applications of process mining techniques in a simulation context are outlined in section 3. For an elaborate overview of process mining, the reader is referred to van der Aalst (2011).

The extracted event log knowledge serves as an additional input for simulation model construction. As mentioned in figure 1, event log knowledge has the ability of complementing and cross-checking traditional information sources. The analysis of event logs might allow the discovery of an alternative activity order, besides the common activity flow derived from traditional information sources. In this way, event log knowledge complements expert interviews, etc. With regards to cross-checking, event data analysis enables the verification of e.g. uncertain statements of business experts concerning particular model elements. Note that a complementing and cross-checking relationship should also exist among the traditional information sources themselves. These relationships are not visualized in figure 1 to maintain its clarity.

An event log can also serve as an initial simulation model construction input. In this respect, the analysis of the event log and initial modeling efforts might e.g. lead to the identification of specific questions that should be directed to a

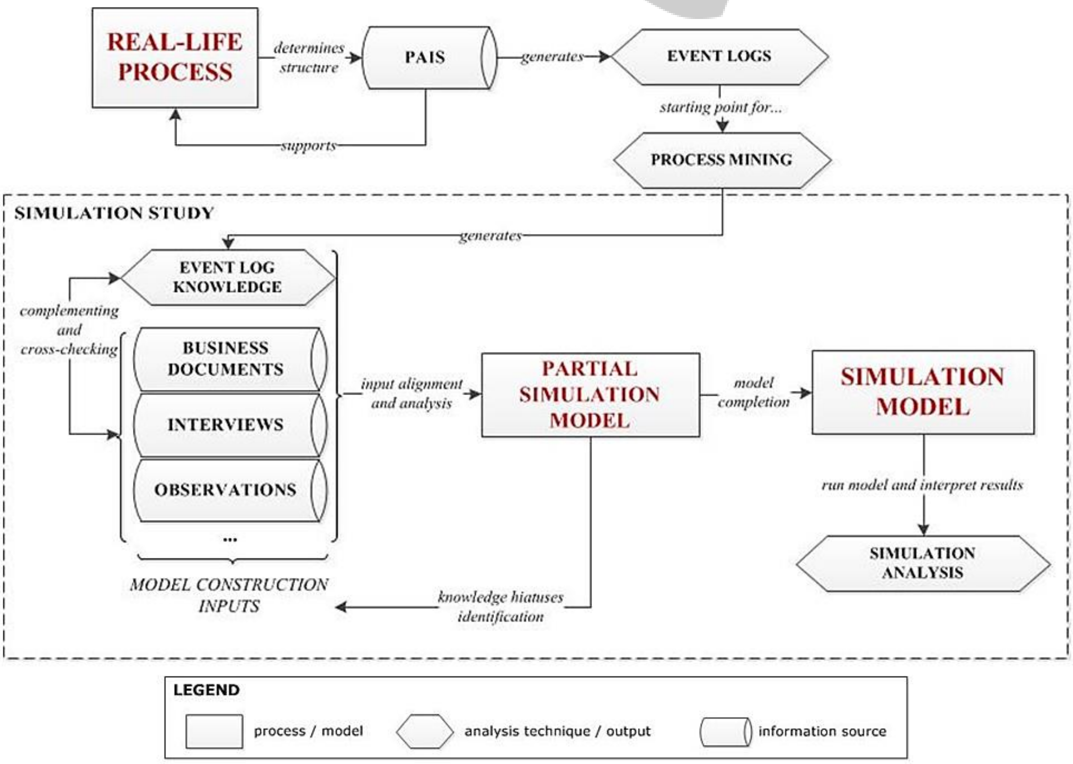


Figure 1: Conceptual framework.

business expert. Consequently, a reiteration occurs from the partial simulation model to the model construction inputs, in this case expert interviews. Besides the identification of appropriate questions, gaining initial process insights using event log analysis can improve the quality of the expert interview's output. This statement is in accordance with Pfadenhauer (2009) as it is deemed important that the interviewer is knowledgeable about the matter of discussion in order to retrieve valuable information of an expert interviewee.

A final remark regarding the conceptual framework relates to the fact that the framework does not advocate for the construction of a simulation model solely based on process execution data due to technical and behavioural issues. On a technical level, process mining outputs might render rather complex results when applied to a real-life process. Consider the use of process mining techniques to discover the activities in a process and their relationships, as will be detailed in section 3.2. When these methods are applied to rather unstructured processes, this might result in a process model with a so called spaghetti structure. The latter is characterized by activities with a large number of mutual interconnections, rendering the process model difficult to understand (van der Aalst, 2011). Besides output complexity, a modeler should also be aware of obstacles such as registration errors or inaccuracies in the log files.

With regards to behavioural issues, building a simulation model solely from an event log might become a black box for business experts and decision makers as they may not be familiar with the used process mining techniques. Moreover, a simulation model can be provided which is out of scope with the personal process insights of the business experts. When this is the case, the acceptance and hence the use of the simulation model might be hindered. The latter is, amongst others, supported by the confirmation bias concept, which refers to the human tendency to ignore or criticize information that contradicts their proper beliefs (Zimbardo et al., 2005).

# 3 EVENT LOG KNOWLEDGE IN SIMULATION MODEL CONSTRUCTION

The previous section presented a conceptual framework advocating the integration of event log knowledge as an additional and complementary input for simulation model construction. The inclusion of event log knowledge was mentioned in rather generic terms. This section provides additional insights in the use of event log knowledge in a simulation model construction setting. Furthermore, preliminary research efforts within this domain are outlined.

A simulation model is composed of several building blocks. Based on a literature review, several generic simulation model components are identified. The discussion in this section is organized using some key simulation model building blocks. Due to space limitations, the present paper will focus on entities and interrelated activities. Nevertheless, event log knowledge can also have a valuable contribution when modeling attribute values, resources and the simulation model's decision logic. The latter two are key points on our research agenda.

## 3.1 Entities

An entity is a dynamic object that is created within the simulation model, moves throughout the organizational system and afterwards typically leaves the system. Entities cannot be considered as a homogeneous set of objects: various entity types can coexist within the model (Law, 2007).

The definition of entity types can be useful as the followed order of activities, activity duration distributions, etc. might differ for each entity type. Moreover, distinguishing entity types allows for the retrieval of decomposed performance statistics as the introduced distinction is typically maintained when performance metrics are generated.

Business experts might find it difficult to define the appropriate number of relevant entity types with respect to the objective of the simulation study. Within this context, event logs can provide useful insights as they contain information on the activity order of a particular entity and possibly the activity duration for this entity. Moreover, entity attributes might be recorded as case attributes. Entities following similar routes in the process or having converging attribute values can be grouped. These clusters can be perceived as possible entity types for the process under consideration and could form a starting point for a discussion with business experts. Despite the potential benefits of process mining in entity type discovery, no research efforts are identified on this matter. Consequently, the identification of entity types is one of the focal points on our research agenda.

## 3.2 Interrelated Activities

Once entities have entered the organizational system, they flow through a series of interrelated activities which provide a particular service to them (Tumay, 1996).

To model process activities and their relationships, the modeler might resort to a traditional information source such as expert interviews. Even though discussions with business experts will provide valuable knowledge regarding interrelated activities modeling, some aspects such as deviating activity orders might not become apparent. This type of process behaviour might however be relevant to include in the simulation model if it has a non-negligible impact on the operation of the process. As event logs contain information on activity orders, log analysis can provide valuable insights in this respect.

Several algorithms have been developed to discover process models from event logs. Some of these techniques have been applied to determine the order of activities in a simulation model; e.g. the alpha-algorithm (Rozinat et al., 2009), heuristic mining (Măruşter and van Beest, 2009) and fuzzy mining (van Beest and Măruşter, 2007). For more technical details on these algorithms, the reader is referred to van der Aalst (2011).

Publications on this matter tend to consider fairly simple processes. This observation can possibly be attributed to the rather conceptual nature of the literature in this domain, where activity order discovery serves as a proof-of-concept for the proposed methodology. As indicated in the final paragraph of section 2.2, process mining algorithms might generate results with a high degree of complexity. The modeler can however also use process discovery techniques to prepare expert interviews or improve the activity order determined through the use of traditional information sources.

# 4 ILLUSTRATIVE SIMULATION STUDY

To demonstrate that the combination of traditional information sources and event log knowledge can be valuable during simulation model construction, this section outlines an illustration based on real-life data. For reasons of confidentiality, the illustration has been anonymized.

The company under consideration provides roadside assistance services, i.e. the company offers its members assistance when confronted with a car or motorcycle breakdown. If the latter occurs, a member can contact the central dispatching center. Consequently, the assistance request is transmitted to a patrolman who is responsible for the repair. To limit the inconvenience for the motorists, the company guarantees its members that for 90% of the requests, a patrolman will be present to help them within 45 minutes.

The roadside assistance company was, amongst others, interested in the effects of a changing workforce size on the operational performance. For this purpose, a simulation model needed to be constructed. During the development of the simulation model, both event log knowledge and expert knowledge were used. The provided event log contained information on the activities of the central dispatching center on the one hand, e.g. the receipt of an assistance request, and the patrolmen on the other hand, e.g. the arrival at the requested location. Expert knowledge was obtained during a meeting with a staff member of the roadside assistance company.

Simulation model construction efforts have shown that event log insights and expert knowledge are complementary. Consider e.g. the determination of input parameters such as service time distributions. The use of event logs allowed for a more accurate parameter estimation than the rather rough approximations a business expert would have been able to provide. Expert input can however be valuable to avoid the use of distorted service time distributions due to the inclusion of extreme outliers caused by registration errors.

Another example showing the complementarity among event log and expert knowledge relates to the activities comprising the process and their relationships. The analysis of the event log highlighted a dominant path that about two thirds of the assistance requests follow: request receipt – request transmission to patrolman – request acceptance by patrolman – patrolman departure – patrolman arrival – assistance termination – return of patrolman. This might also be the business experts' response when requested to describe the process. However, other activity orders with a less straightforward interpretation could be identified from the event log. Some of these activity patterns might reflect relevant process behaviour which should be included in the simulation model, while other activity orders are a consequence of registration errors or inaccuracies. By engaging in a dialogue with the business expert, the modeler could determine which patterns had to be included in the simulation model. Adding additional relevant patterns will enhance the realism of the simulation model and consequently the usefulness of the analysis results.

The examples in the previous paragraphs show that event log knowledge can provide valuable additional insights that can be integrated in the simulation model. Moreover, event log analysis can allow the modeler to identify relevant questions and discussion topics which should be directed to the business expert. In this sense, traditional information sources and event log knowledge should not be seen as adversaries but as complementary model construction inputs.

# 5 DISCUSSION AND CONCLUSIONS

Business process simulation models are typically created based on insights from traditional information sources such as process documentation, interviews and observations. Issues with these model construction inputs contribute to the discrepancy between the constructed simulation model and reality (Rozinat et al., 2009). The use of process execution data, representing real-life process behaviour, can be a valuable instrument to bridge the gap between model and reality to a certain extent.

This paper advocates the use of event log knowledge as an additional and complementary simulation model construction input. Event log knowledge is obtained by applying process mining techniques to event logs, i.e. files capturing process execution data which are retrieved from a PAIS. A conceptual simulation model construction framework is presented which includes this additional input. Moreover, the discussion of the framework and the illustration in section 4 highlight the complementarity of event log knowledge and traditional information sources.

The integration of event log knowledge in simulation model construction is mentioned in rather generic terms in the conceptual framework. Some of the scarce publications regarding the use of process mining in a simulation context have been outlined. Our research agenda includes further efforts to benefit from event log knowledge with regards to the modeling of entity types, decision logic and resources. Specific domains in which new techniques are planned to be developed or existing research will be leveraged includes entity type discovery, the determination of priority rules in queues and the modeling of human resources behaviour.

It should be noted that current publications on the use of process mining in a simulation context tend to limit themselves to state how process mining can be useful when constructing a simulation model. However, they typically fail to recognize that event log analysis results can be overly complex or contain inconsistencies due to registration errors. Consequently, the obtained event log knowledge needs to be cross-checked with business experts.

This paper explicitly states that event log knowledge is complementary to traditional information sources and that both can strengthen each other. In this way, a more realistic, but manageable, simulation model can be created compared to a situation in which traditional information sources or event log knowledge are used in isolation. An improvement of the simulation model realism will lead to more representative analysis results and hence more adequate decision support.



# REFERENCES


Brysbaert, M. (2006). *Psychologie*. Gent: Academia Press.

Law, A.M. (2007). *Simulation modeling & analysis.* New York: McGraw-Hill.

Măruşter, L. and van Beest, N.R.T.P. (2009). Redesigning business processes: a methodology based on simulation and process mining techniques. *Knowledge and Information Systems*, 21(3), 267-297.

Melão, N. and Pidd, M. (2003). Use of business process simulation: a survey of practitioners. *Journal of the Operational Research Society*, 54(1), 2-10.

Pfadenhauer, M. (2009). At eye level: the expert interview - a talk between expert and quasi-expert. In Bogner, A., Beate, L., Menz, W., (Eds.). *Interviewing experts*. New York : Palgrave Macmillan, pp. 81-97.

Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences*, 11(1), 37-43.

Rozinat, A., Mans, R. S., Song, M. and van der Aalst, W.M.P. (2009). Discovering simulation models. *Information Systems*, 34(3), 305–327.

Rozinat, A., Wynn, M.T., van der Aalst, W.M.P., ter Hofstede, A. and Fidge, C.J. (2008). Workflow simulation for operational decision support using design, historic and state information. *Lecture Notes in Computer Science*, 5240, 196-211.

Tumay, K. (1996). Business process simulation. *Proceedings of the 1996 Winter Simulation Conference*, Coronado, USA, December 1996, 55-60.

Van Beest, N.R.T.P. and Măruşter, L. (2007). A process mining approach to redesign business processes – a case study in gas industry. *Ninth International*

*Symposium on Symbolic and Numeric Algorithms for Scientific Computing*, 541-548.
Van der Aalst, W.M.P. (2011). *Process mining: discovery, conformance and enhancement of business processes*. Heidelberg: Springer-Verlag.
Vincent, S. (1998). Input data analysis. In Banks, J. (Ed.), *Handbook of simulation: principles, methodology, advances, applications and practice*. Hoboken: John Wiley & Sons, pp. 3-30.
Zimbardo, P.G., Weber, A.L., and Johnson, R.L. (2005). *Psychologie: een inleiding* (S. Birnie, Trans.). Amsterdam: Pearson Education. (Original work published 2003).