

PMSD: Data-Driven Simulation Using System Dynamics and Process Mining ^{*}

Mahsa Pourbafrani and Wil M. P. van der Aalst

Chair of Process and Data Science, RWTH Aachen University, Germany
{mahsa.bafrani,wvdaalst}@pads.rwth-aachen.de

Abstract. Process mining extends far beyond process discovery and conformance checking, and also provides techniques for bottleneck analysis and organizational mining. However, these techniques are mostly *backward-looking*. *PMSD* is a web application tool that supports forward-looking simulation techniques. It transforms the event data and process mining results into a simulation model which can be executed and validated. *PMSD* includes log transformation, time window selection, relation detection, interactive model generation, simulating and validating the models in the form of system dynamics, i.e., a technique for an aggregated simulation. The results of the modules are visualized in the tool for a better interpretation.

Keywords: Process mining · Simulation · System Dynamics · What-if analysis

1 Introduction

Process mining uses stored event data of organizations, i.e., event logs, to provide actionable insights for organizations [1]. Different tools address process discovery, performance analysis, bottleneck analysis, and deviation detection. Yet, the gap between the *backward-looking* and the *forward-looking* process mining techniques remains. Traditional forward-looking techniques as mentioned in [2], use events in the process as a basis of simulation. They aimed to mimic the process at the level of detail and simulate it. In more recent simulation tool such as [3], different level of detail for simulation is acquired, e.g., duration of activities and the flow of activities are used. Moreover, the *Monte Carlo* technique is used in the *pm4py* tool¹ for simulating discovered Petri nets.

In *PMSD*, we use the idea that a simulation model can be learned from the event data at an aggregated level. The traditional connections between process mining and simulation mainly use a descriptive model discovered in the discovery step to enrich the simulation models at the level of the process instances, e.g., Discrete Event Simulation (DES). The presented tool is the result of our

^{*} Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2023 Internet of Production- Project ID: 390621612. We also thank the Alexander von Humboldt (AvH) Stiftung for supporting our research.

¹ <http://pm4py.pads.rwth-aachen.de>

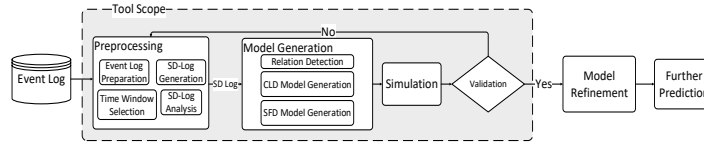


Fig. 1. Our proposed framework for using process mining and system dynamics together in order to design valid models to support scenario-based prediction of business processes in [4]. This paper focuses on the *developed tool*, i.e., the highlighted step.

approach in generating simulation results for business processes at an aggregated level providing the option to add external factors into the simulation [4]. Figure 1 shows the overview of the approach starting from an event log and ending with a scenario-based simulation model. The steps indicated in the highlighted parts are supported by the tool. We extract possible variables from the process in different steps of time instead of taking the events into account for the simulation as shown in Fig. 2.

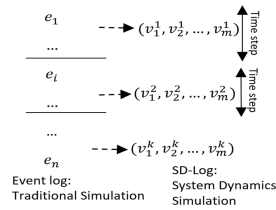


Fig. 2. Traditional Simulation vs. *PMSD*. We extract possible variables (m) over time steps (k).

variable notations. Flows add/remove to/from the values of stocks, also, variables affect/get affected by the flows, other variables. *PMSD* provides insights through the processes over time which can be hidden from the user, e.g., a non-linear relation between the workload of resources and the speed of performing tasks.

2 Description of Functionalities

In our approach, the possible process variables are extracted over time, e.g., arrival rate per day and average service time per day. The newly generated log (SD-Log) is the cornerstone of the simulation. The preprocessing step and extracting the best parameters in the framework by means of time series analysis

The Model generation module is introduced in [6] and the preprocessing step is presented in [5]. The event log is transformed into a set of variables over time and the values of these variables form the System Dynamics logs (SD-Logs). To generate more stable SD-Logs, we use time series analysis over the values. The relations between variables over time in the SD-Log are used for creating the system dynamics models. We support both causal loop diagrams (CLD) and stock-flow diagrams (SFD). System dynamics models the systems and their relations with the environment [8]. CLDs represent these conceptual relationships and SFDs model the underlying equations using *stock*, *flow* and

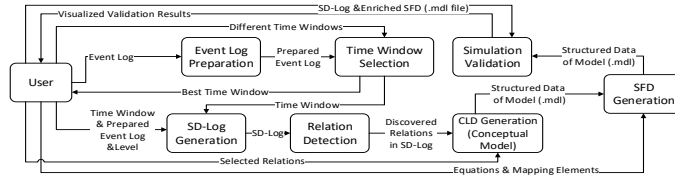


Fig. 3. Data flow diagram of the *PMSD* including data flow between the user and the main modules as well as the background flow of data between the modules.

proposed in [5]. To form a valid system dynamics model, we have to discover all the relations, i.e., linear and nonlinear correlations, between the generated process variables over time as introduced in [6]. Analyzing a process and creating aggregated features of the process over time (process variables) for further analyses is the main focus of the tool.

PMSD is being designed in such a way that in all the steps, the outputs are accessible for users. Figure 3 depicts the data flow diagram of the application. The inputs and generated outputs in each module and the interactions with the user are shown. The generated SD-Logs including active steps in the processes as well as all the steps for the different selected time windows in the form of *.csv* are captured. Also, all the designed CLDs and SFDs in the *.mdl* format are stored locally for the user. To run the tool locally, the home page can be accessed via any browser using the *http://127.0.0.1:5000* URL. All the modules are designed as different tabs and are visually accessible. *PMSD* is a fully interactive tool with a user interface based on *Python* and *Flask* technology. The results of the steps are shown graphically to provide an easier interpretation possible. It contains 8 tabs and each tab can be run separately with different inputs/output of the other modules/tabs. Currently, the following components are available:

- *Event log transformation* indicates the main attributes of the event log, discovers the directly follows graph, and presents the event log’s information.
- *Time window selection* assesses the quality of the user’s preference for selecting a time window for generating simulation data.
- *Simulation log generator* uses the transformed event log and the selected time window to generate simulation data (SD-Log). It generates an SD-Log for different aspects and levels, i.e., general process, organizational, and activity aspects. For instance, an SD-Log of the general aspect of a process includes the arrival rate of the process, and average service time in the process and other possible measurable variables per day.
- *Relation detection* investigates whether there is any strong relationship between the variables in the extracted SD-Log. Furthermore, the user can look for the relations between variables in different steps of time.
- *Detailed relations*, presents the existing relations between every two variables in the SD-log for further investigation on the types of relations.

- *Interactive conceptual model generation* provides the option for the user to choose between all the strong relations discovered in the relation detection module and creates CLD, i.e., effects and relations between process variables. It generates both the graphical model in the tool and the *.mdl* (text format) file to be used in most of the system dynamics tools, e.g., *Vensim*².
- *Interactive stock-flow diagram* generates SFDs graphically in *PMSD* and the (*.mdl*) file. The relations are directly transformed from the CLD (previous step) and the user can map the process variables to the SFD elements.
- *Simulation and validation* simulates the SFD model using the values in the SD-Log and validates the results using the pair-wise comparison of the SD-Log and simulation results values and their distributions.

3 Maturity of the Tool

The evaluation results of our proposed forward-looking approaches in process mining are represented using different modules of the tool. *PMSD* along with a tutorial and a screen-cast is available on GitHub.³ It has also been used in some industrial projects, e.g., in the project of *Internet of Production* in the context of *Industry 4.0*. In [7], part of the results of using *PMSD* for the production line is presented. By an example, i.e., an event log of a call center designed by the CPN tool, we show some similar results.

Stability Test of Time Window

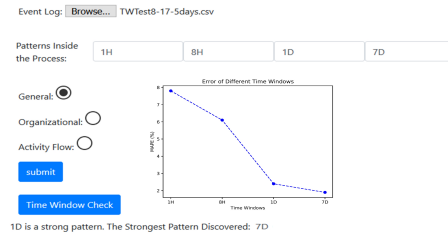


Fig. 4. *Stability test* showing the error of training models for the time windows.

are shown in validation module. The results include a comparison between the real values and the simulated ones and their distributions for the selected variables.

4 Conclusion

In this paper, we introduced *PMSD* to support designing system dynamics models for simulation in the context of business processes. Using *PMSD*, we look into

² www.vensim.com

³ <https://github.com/mbafrani/PMSD>

We use different suggested time windows to extract values over time for the possible process variables using the time window test. The result in Fig. 4 shows the selected time windows by the user and the errors of trained models for each time window. Figure 5 represents the user interface for selecting the strong detected relations between the variables. Finally, by uploading the generated SFD and SD-log (both are automatically generated), the automatic simulation is performed and the validation results

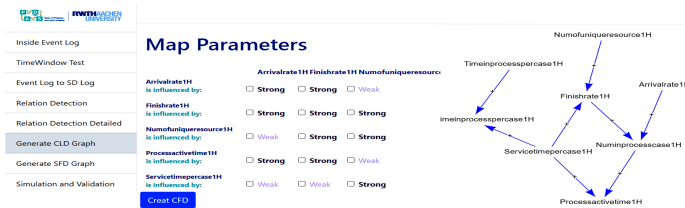


Fig. 5. The *conceptual modeling* section showing the detected relations and their strength between the variables. The user is able to select among the selected relations.

the processes at different aggregation levels, e.g., hourly or daily, as well as different aspects, e.g., overall process or organizational aspects. The provided user interface and the graphical outputs make the interpretation of the results easy. Applying *PMSD*, the underlying effects and relations at the instance level can be detected and modeled in an aggregated manner. Besides the option to simulate and validate the models directly in the tool, the models can be simulated or refined by adding external variables using simulation software like *Vensim*.

References

1. van der Aalst, W.M.P.: *Process Mining - Data Science in Action*, Second Edition. Springer (2016)
2. van der Aalst, W.M.P.: *Process Mining and Simulation: A Match Made in Heaven!* In: *Computer Simulation Conference*. pp. 1–12. ACM Press (2018)
3. Camargo, M., Dumas, M., Rojas, O.G.: *Simod: A tool for automated discovery of business process simulation models*. In: *Proceedings of Demonstration Track at BPM 2019*. pp. 139–143 (2019)
4. Pourbafrani, M., van Zelst, S.J., van der Aalst, W.M.P.: *Scenario-based prediction of business processes using system dynamics*. In: *On the Move to Meaningful Internet Systems: OTM 2019 Conferences - Confederated International Conferences: CoopIS, ODBASE, C&TC 2019, Rhodes, Greece, October 21-25, 2019, Proceedings*. pp. 422–439 (2019). https://doi.org/10.1007/978-3-030-33246-4_27, https://doi.org/10.1007/978-3-030-33246-4_27
5. Pourbafrani, M., van Zelst, S.J., van der Aalst, W.M.P.: *Semi-automated time-granularity detection for data-driven simulation using process mining and system dynamics*. In: *Conceptual Modeling - 39th International Conference, ER 2020, Vienna, Austria, November 3-6, 2020, Proceedings* (2020)
6. Pourbafrani, M., van Zelst, S.J., van der Aalst, W.M.P.: *Supporting automatic system dynamics model generation for simulation in the context of process mining*. In: *Business Information Systems - 23st International Conference, BIS 2020, Colorado Springs, USA, 8-10 June, 2020, Proceedings* (2020)
7. Pourbafrani, M., van Zelst, S.J., van der Aalst, W.M.P.: *Supporting decisions in production line processes by combining process mining and system dynamics*. In: *Proceedings of the 3rd International Conference on Intelligent Human Systems Integration*. pp. 461–467 (2020). https://doi.org/10.1007/978-3-030-39512-4_72
8. Sterman, J.: *System Dynamics: Systems Thinking and Modeling for a Complex World* (2002)