A Methodology for Aligning Process Model Abstraction Levels and Stakeholder Needs

Dennis G. J. C. Maneschijn, Rob H. Bemthuis, Faiza A. Bukhsh, and Maria-Eugenia Iacob
University of Twente, Drienerlolaan 5, 7522 NB, Enschede, The Netherlands

Keywords: Process Mining, Abstraction Levels, Stakeholder Analysis, Process Models.

Abstract: Process mining derives knowledge of the execution of processes through analyzing behavior as observed from real-life events. While benefits of process mining are widely acknowledged, finding an adequate level of detail at which a mined process model is suitable for a specific stakeholder is still an ongoing challenge. Process models can be mined at different levels of abstraction, often resulting in either highly complex or highly abstract process models. This may have an important impact on the comprehensibility of the process model, which can also differ from the perspective of a particular stakeholder. To address this problem from a stakeholder-centric perspective, we propose a methodology for determining an appropriate level of process model abstraction. To this end, we use quantitative metrics on process models as well as a qualitative evaluation by using a technology acceptance model (TAM). A logistics case study involving the fuzzy process mining discovery algorithm shows initial evidence that the use of appropriate abstraction levels is key when considering the needs of various stakeholders.

1 INTRODUCTION

Process mining helps to unveil actionable knowledge and insights of a process, based on historical execution data in the form of event logs (Van der Aalst, 2016). Process mining aims to discover (i.e., learning a model from behavior recorded in an event log), monitor, and improve processes based on event logs. Process mining, as a relatively young discipline, could provide valuable insights in, for example, performance evaluation, root-cause analysis, bottleneck analysis, and process prediction and optimization (Van der Aalst, 2016).

Although process mining has the potential to provide many promising insights and to support process improvements, there are still many challenges to overcome (Dos Santos Garcia et al., 2019). For example, research on bottleneck analysis techniques utilizing process mining is still an under-researched area (Bemthuis et al., 2021b). Another challenge concerns finding an appropriate balance between a process model that is comprehensible, but still sufficiently detailed in order to show relevant behavior to the user of the model (Van der Aalst and Gunther, 2007; Leemans et al., 2020). Sometimes, the generated process models tend to be confusing and difficult to understand, especially when there are many diverse cases with deviating behaviors. Although this complexity can be useful for some stakeholders inside an organization, generally speaking, applying no abstraction could make the model too complex to comprehend (Van der Aalst, 2016).

To deal with process model complexity, abstraction can be applied, which can make the process model less spaghetti-like and more comprehensible. Abstraction omits lower-level information, which is insignificant in the chosen context, from the visualization (Van der Aalst and Gunther, 2007). For example, typically not all small roads and pedestrian paths are shown to bus drivers when considering a city road map since it would make the map cluttered and inconvenient to use. We consider the following definition of abstraction: "Simplifying process models by removing edges, clustering nodes, and removing nodes to make the process model more suitable for the person looking at it." In other words, abstraction is about the level of granularity of the process model.

A model can have a too high abstraction (underfitted) or a too low abstraction (overfitted) (Van der Aalst and Gunther, 2007). Finding an appropriate level of abstraction also depends on which user/stakeholder is using the model, because users typically have different needs and purposes. Yet, providing a process model that is suitable for a des-

Maneschijn, D., Bemthuis, R., Bukhsh, F. and Iacob, M.
A Methodology for Aligning Process Model Abstraction Levels and Stakeholder Needs.
DOI: 10.5220/0011029600003179
In *Proceedings of the 24th International Conference on Enterprise Information Systems (ICEIS 2022) - Volume 1*, pages 137-147
ISBN: 978-989-758-569-2; ISSN: 2184-4992
Copyright © 2022 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

ignated stakeholder can be challenging. For example, in the logistics domain where many stakeholders have different needs, experiences, and backgrounds (Flodén and Woxenius, 2021; Tolentino-Zondervan et al., 2021). Real-world processes often involve unstructured and ad-hoc behavior, which produce spaghetti-like models. It seems reasonable to argue that simplification and abstraction are needed at higher levels of management, but especially those working 'in the trenches' of a process not only must know about the details, they can often also tell you why those details are there and whether or not these steps are needed or can be circumvented/re-engineered.

Multiple papers attempt to address the issue of abstraction in process mining (Günther and Van Der Aalst, 2007; Baier et al., 2014; Kumar et al., 2017; Van Cruchten and Weigand, 2018b). However, to our knowledge, none of these works try to seek an abstraction level that serves the purpose of a particular stakeholder. Instead of applying abstraction on process models directly (e.g., through filtering), abstraction has also been applied to the level of event data. A literature review on event abstraction already showed the importance of applying pre-processing techniques for the successful application of process mining (Van Zelst et al., 2021), such as in large-scale industrial ERP systems. Nevertheless, as mentioned by these authors, as well as concluded by research on classifying event abstraction articles (Diba et al., 2020), today's approaches still often rely on strong assumptions and domain experts.

We acknowledge the importance of applying abstraction at the level of pre-processing, but significant benefits can also be obtained once a process model has been created, because sometimes it might not even be possible to apply pre-processing techniques on event records (for example, when dealing with noisy or incomplete data sets) (Zakarija et al., 2015; Van Zelst et al., 2021). Alternatively, it could be the case that we only have the process model distilled from many events and that we do not, or only to a limited extent, have information about the exact event records that were used to generate the process model. For example, due to privacy concerns, only some of the original event log's content may be revealed (Fazzinga et al., 2018b). Hence, it may not be possible anymore to gather the original event logs.

In the realm of process model complexity, parallels may be drawn with the discipline of enterprise architecture (EA). Many stakeholders involved in EA have a different perception of the complexity of an architectural model (Iacob et al., 2018). Furthermore, EA models are often showcased from a particular viewpoint which could be specifically designed for a target group. For example, business executives may focus on its value delivery, management on its functionalities and costs, IT architects on its maintainability, and software developers on its flexibility (Iacob et al., 2018). In turn, these different perceptions may lead to disagreement and mismanagement. Similar reasoning has been mentioned regarding business process model complexity metrics (Gruhn and Laue, 2007; Muketha et al., 2010). The notion of using both objective and subjective metrics for striving towards an optimal level of complexity to effectively and efficiently understand and use process mining, is an area we aim to further explore.

To summarize, the benefits of using process mining are widely acknowledged, but the generated process models can be challenging to understand for (business) users (Yazdi et al., 2021). Therefore, we may not achieve the (intended) goals of process mining because we are unable to review the as-is process models in comparison to the to-be process model (to a sufficient degree). To address this shortcoming, this paper proposes a methodology that deals with model complexity, and aligns it with the needs of various stakeholders. We focus on the application of abstraction within the scope of the discovery phase of process mining, as discussed in the process mining project methodology of (Van Eck et al., 2015). Since our main research product (i.e., design artifact) is a methodology, we followed Peffer's Design Science Research Methodology (DSRM) (Peffers et al., 2007) during this research study.

The remainder of this paper is structured as follows. Section 2 presents the proposed methodology followed by Section 3 where the methodology is demonstrated by using a logistics case study. Section 4 discusses the related work. Finally, Section 5 concludes and gives some pointers to future work.

2 A METHODOLOGY FOR ALIGNING PROCESS MODEL COMPLEXITY AND STAKEHOLDER NEEDS

The proposed methodology in this paper deals with the problem of making process models comprehensible for organizational stakeholders. The methodology consists of the six phases as shown in Figure 1, each requiring a certain input and delivering an output. Our methodology is inspired by CRISP-DM (Wirth and Hipp, 2000), a well-known methodology for data science projects. In the remainder of this section, we briefly explain each phase.

Phase 1 - Business Understanding & Data Preparation. The first phase forms the basis for conducting the other phases of our methodology. It allows the user to get familiar with the domain in which the research is conducted. Furthermore, it requires the user to get to know the data to be worked with and which preparation steps were conducted.

Phase 2 - Identifying Relevant Experts/ Stakeholders. The second phase uses the business understanding to define the relevant experts/stakeholders for which one wants to determine the appropriateness concerning the generated process models. Apart from defining the relevant experts/stakeholders, it is also important to identify the (business) goals of each expert/stakeholder.

Phase 3 - Establishing Abstraction Levels. Using the stakeholders and business purposes from phase 2, phase 3 defines the different abstraction levels at which we aim to produce the process models. This can be done by considering the characteristics of the chosen process miner, in combination with the data understanding and expert/stakeholder overview. The goal is to have a distinct separation of organizational levels represented through specified abstraction levels.

Phase 4 - Defining Quality Measurements. Before generating the process models, it is important to determine the quality measures that will be used in the evaluation phase. It is recommended to determine both quantitative and qualitative measures. In the end, our methodology produces a social impact since it attempts to find a solution to a model complexity problem faced by stakeholders. Therefore, we highlight the importance of using a quantitative and qualitative measuring tool.

Phase 5 - Modeling. In this phase, we select and apply an algorithm for process model extraction.

Phase 6 - Evaluation. The final phase uses the stakeholder overview, quality measurements, and process models to evaluate the quality of the process models for each stakeholder using an adaptation of the Technology Acceptance Model (TAM) (Davis, 1989). In the end, the evaluation provides an alignment of stakeholder needs with process model complexity.

3 DEMONSTRATION

The case study used to demonstrate our methodology is introduced in earlier work (Bemthuis et al., 2020). This study concerns the transport of perishable goods in a production facility. Multiple data sources, such as smart pallets, are used to collect data about the state of the shipments, transport units, and the environment (Bemthuis et al., 2020). The production facility strives towards a good balance between minimizing quality decay against minimizing operating costs. Investigating this case study seems promising as multiple stakeholder perspectives could be considered as well as complexity levels, because of the public data set (Bemthuis et al., 2021a) and because many of the partners involved in the previous work are also involved in the present research. This section will apply each phase of our methodology to this case study.

3.1 Business Understanding & Data Preparation

The business understanding mainly included understanding the case study and getting to know characteristics of the logistics domain, as well as the challenges this brings forward in process mining. Logistics processes are known for being complex and dynamic, often producing spaghetti process models, which seems an interesting domain for our study.

The data preparation phase consists of enriching the event logs of the case study. The event logs describe activities about movable transport units that transport smart pallets from one place to the other. The pallets and transporters are equipped with sensors that keep track of the status of products and/or transporters (Bemthuis et al., 2020). The quality of products decays over time. How fast a product quality is depreciating, depends on, e.g., the type of vehicle and type of food transported. The data set comprised several scenarios and experimental runs per scenario (Bemthuis et al., 2021a). We considered the scenario and experimental results of which the average product decay was the lowest and of which a warm-up period was removed. More details about the case study can be found in (Bemthuis et al., 2020).

Two additional attributes of the event log were used: the decay level (DL) of a product and a vehicle identifier. The DL is recorded as a numerical value. To make the stakeholder assessment comprehensible for the stakeholders, we classified the DL into four categories. The categories can be found in Table 1. These categories are based on the mean and the standard deviation of the DL per scenario. We have decided to not use predefined quality levels, because we aimed to obtain events that are fairly distributed over the different quality levels (i.e., proportionally for the scenario). After enriching the data with this categorical value, we filtered the event log to remove incomplete traces. Filtering was done using the heuristics filter plug-in of the open-source tool ProM.

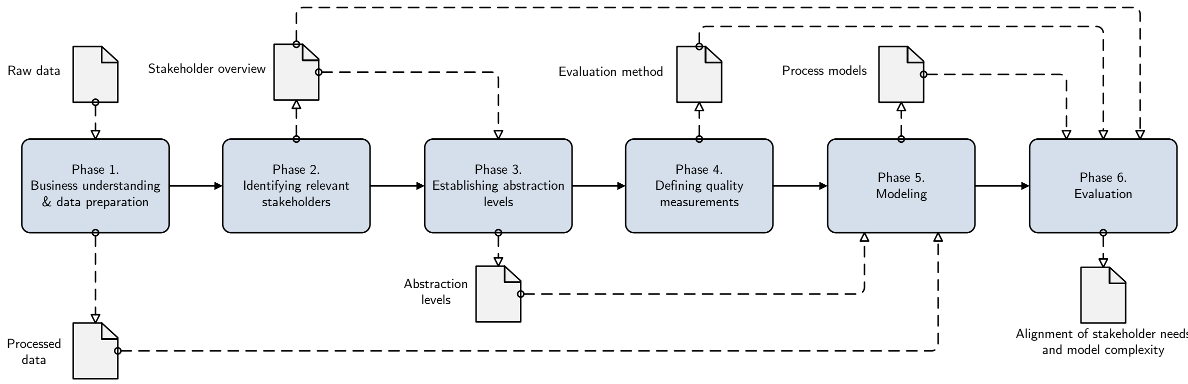


Figure 1: Methodology followed for aligning stakeholders to process model abstraction levels.

Table 1: Product quality decay categories used for data enrichment.

Quality category	Partition of
Good	$DL \geq \mu + \sigma$
Sufficient	$\mu \leq DL < \mu + \sigma$
Insufficient	$\mu - \sigma < DL < \mu$
Poor	$DL \leq \mu - \sigma$

3.2 Identifying Relevant Stakeholders

Actively managing stakeholders and addressing the needs of stakeholders is beneficial for an organization (Greenley and Foxall, 1997; Post et al., 2002). As mentioned before, process mining can improve business processes. However, the usefulness of mined process models depends on, e.g., whether the model is comprehensible for the stakeholder or not. It is interesting to see how different abstraction levels influence the appropriateness of the process model for a specific stakeholder. We will determine a list of stakeholders of a logistics organization and define the organizational level of interest for each stakeholder. These interest levels will form the basis of the abstraction levels as defined in phase 3 of our methodology.

To this end, multiple stakeholders inside logistics organizations were considered. Such organizations typically include both primary and secondary stakeholders (Flodén and Woxenius, 2021; Iacob et al., 2019; Tolentino-Zondervan et al., 2021). Primary stakeholders have a formal, or contractual relationship to the organization, while secondary stakeholders are not directly connected to the company (Gibson, 2000). Since secondary stakeholders are typically not directly bothered with analyzing a process model, we will not put them into our stakeholder list. Instead, we will define primary stakeholders that represent the needs of secondary stakeholders.

Stakeholders usually have different goals and interests. Furthermore, stakeholders are generally concerned with different kinds of information. It might be that one particular stakeholder wants to know more about the overall structure of the process, while another stakeholder is more interested in specific (abnormal) activities. For each stakeholder, we will define a general purpose that is aligned with his needs. By considering the case study context, work of (Anthony, 1965; Greenley and Foxall, 1997; Iacob et al., 2019; Post et al., 2002) helped us to identify the goals of the stakeholders. The list of stakeholders considered is shown in Table 2.

3.3 Establishing Abstraction Levels

The fuzzy miner is used for generating the process models. This miner uses a combination of significance and correlation thresholds to simplify the resulting model (Van der Aalst and Gunther, 2007; Günther and Van Der Aalst, 2007). Significance is about the relative importance of behavior, while correlation is about the precedence relation of two events. Being able to tune on these parameters, makes the fuzzy miner a suitable candidate for handling complex and unstructured real-life event logs.

The parameters of the fuzzy miner have different influences on the obtained process model. Five thresholds can help to simplify the process model. Two of these thresholds, the preserve threshold and the ratio threshold, will not be considered, because these only have an influence on nodes with a conflicting relation (Günther and Van Der Aalst, 2007). Our data set does not contain these types of relations. The other three thresholds all influence the simplification in different ways. An overview of the thresholds can be found in Table 3.

In general, the lower the defined thresholds are, the less abstract a process model will become. The utility ratio, however, does not directly influence the abstraction of the model since it focuses on a combination of significance and correlation. Therefore, we

Table 2: An example of stakeholders with organizational abstraction levels.

Stakeholder	Purpose	Organizational interest
Operational board	Identify the overall workflow of the company	Top
CFO	Get to know the overall cost picture and identifying the specific causes of high costs	Top & Middle
Planner	Identify all the steps an order undergoes and where delays occur	Middle
Driver	Find out what activities constitute to their specific task	Bottom
Exception manager	Spot exceptions and find out how they occurred	Bottom
IT expert	Find out what parts of the process require more extensive logging	Middle
Regulations expert	Make sure all steps necessary for regulation measures are taken	Bottom
Customer relations	Ensure traceability and timeliness of the orders	Bottom

Table 3: Considered thresholds for the fuzzy miner.

Threshold	Value = 0	Value = 1	Application
Utility ratio (UR)	High correlation/ low significance	High significance/ low correlation	Edge filtering
Edge cutoff (EC)	Diminishes utility ratio	Amplifies utility ratio	Edge filtering
Node cutoff (NC)	Less abstract	More abstract	Node filtering

will keep this value constant (to make sure it will not bias our results). Due to the enrichment of the data, there are not many activities with a high significance. Therefore, we kept the node cutoff at a relatively low level, to prevent the model from containing only one cluster.

In total, we consider four abstraction levels (A, B, C, D) (see Table 4), based on discussions among the authors of this paper. Abstraction level A is the most abstract and, therefore, contains the least details. Abstraction level D is the least abstract. We consider model A to correspond to the top organizational level, models B and C with the middle organization level, and model D with the bottom level.

Table 4: Abstraction levels.

Abstraction level	UR	EC	NC
A	0.5	1.0	0.4
B	0.5	0.8	0.25
C	0.5	0.6	0.1
D	0.5	0.4	0.0

3.4 Defining Quality Measurements

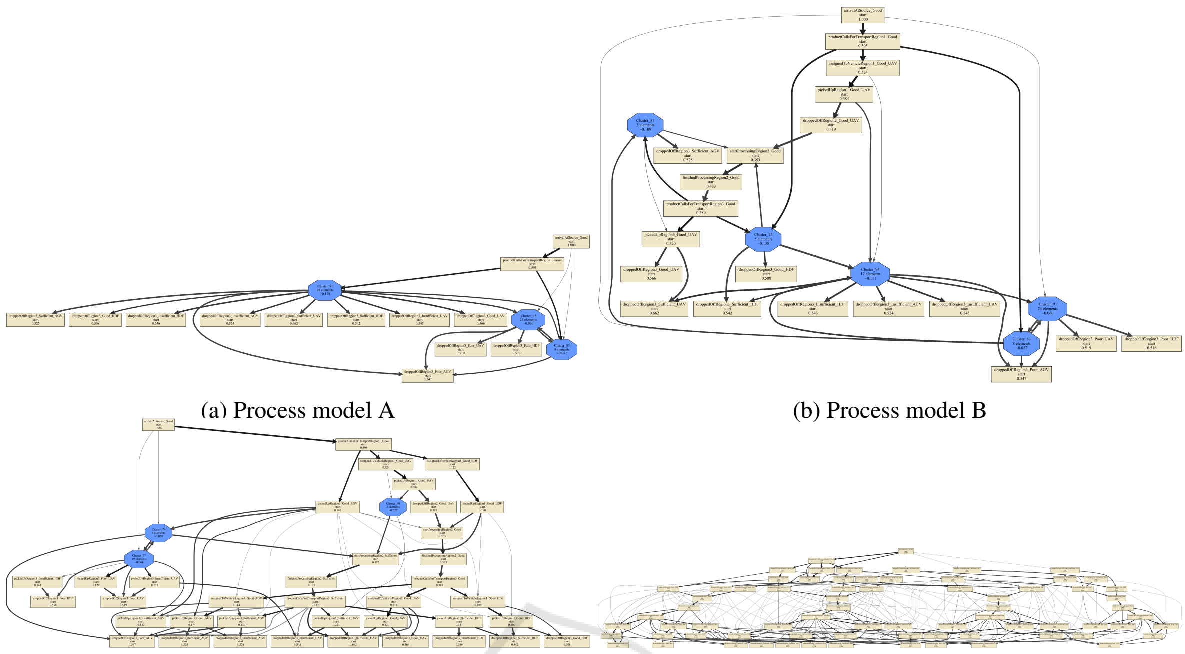
Fitness is an important quantitative measure in process mining, which indicates how well the behavior as described in an event log is displayed in the process model (Van der Aalst et al., 2006). Fitness is useful for getting an idea of the quality of the process model. If we know that such a model is not properly representing reality, we could also be less interested in some other details of the process model. Apart from the fitness, several statistics of the process model are used to obtain an understanding of the complexity of the process models. The first statistic is the level of detail (a percentage that displays how many nodes are preserved in the model). The other statistics include the number of nodes, edges, and clusters shown in the model.

Besides these quantitative measures, we also include a qualitative assessment by using the TAM in combination with an expert analysis. The analysis consists of a panel of stakeholders/experts that evaluate the process models, while reasoning from the perspective of a certain stakeholder. TAM offers a set of questions about the perceived usefulness and ease of use for an end-user. TAM is a well-known approach for measuring the acceptance of new technology, and has already been adopted by researchers within the context of process mining projects (Wynn et al., 2017; Graafmans et al., 2021). We will adapt the original TAM in order to encapsulate the views of stakeholders differ per abstraction level. This adaption makes the TAM suitable for our research to reason on the quality of the process models by directly taking into account the opinions of relevant stakeholders, as defined in Section 3.2. Besides our own research, the proposed TAM can be used in other process mining research, that actively involves stakeholder opinions.

These quality measurements will be used to reason on the quality of a process model. We define quality as the appropriateness of a process model in terms of its fitness, perceived usefulness, and perceived ease of use.

3.5 Modeling

The modeling phase is where the process models will be specified at the abstraction levels (as defined in Section 3.3). The resulting process models are shown in Figure 2.



(a) Process model A (b) Process model B

(c) Process model C (d) Process model D

Figure 2: Process model abstractions.

3.6 Evaluation

The final phase of our methodology is about the evaluation of the generated process models. The goal of the evaluation is to define which level of abstraction is the most useful for a particular stakeholder. By using the measurements defined in phase 4 (Section 3.4) we have performed both quantitative and qualitative evaluation.

3.6.1 Quantitative Results

Our quantitative analysis (see Table 5) shows that, once we get to the more abstract levels, fewer nodes are present and more clusters exist. Model D, the most detailed model, includes all activities and many edges. This causes the fitness of the model to be high, since many specific traces are visible in model D. Although these specific traces can give interesting insights, it might be the case that the model is perceived as cluttered. Model C already places certain nodes inside clusters and heavily reduces the number of edges. Unfortunately, the model has a low fitness score. An explanation can be that many edges have been removed and that not that many nodes are placed in clusters. Model B puts even more nodes in clusters and reduces the number of edges even further. The fact that the clusters are relatively small might improve comprehensibility, since large clusters can be perceived as a black box. Although its fitness is not as high as models A and D, we still consider the fitness to be of a sufficient level. Finally, model A provides the least detail by removing quite some edges and putting almost all nodes inside clusters. This causes the fitness to be high, however, it might be the case that the model is perceived as incomprehensible due to this lack of detail.

3.6.2 Qualitative Results

The results of the expert/stakeholder analysis in combination with the TAM can be seen in Table 6. The table shows how each model performs in terms of usefulness and ease of use, as perceived by the different stakeholders. In total, eight domain experts were consulted for the expert analysis. The experts included mainly academics active in the logistics domain. Every expert reasoned from the perspective of two stakeholders. This means that every stakeholder is reviewed twice, by two different experts. Each expert gives a score ranging from 1 to 5 on a process model, while reasoning from the perspective of a particular stakeholder. Hence, every process model was reviewed sixteen times, from the view of eight stakeholders in total.

When observing the results (as summarized in Table 7), we see that model A scores worst in terms of usefulness. As mentioned in the quantitative analysis part, due to the lack of detail, the model does probably not provide relevant information. The other mod-

Table 5: Evaluation of process models.

Model	Fitness	Detail	Nodes	Edges	Clusters (nodes)
A	99.84%	52.44%	13	21	$3(28,24,8)$
B	96.26%	69.17%	20	40	$5(24,12,8,5,3)$
C	91.99%	88.71%	38	93	$3(19,6,2)$
D	99.00%	100%	73	150+	0

Table 6: Results Technology Acceptance Model (scale 1-5).

Construct	**Average**			
General Information				
1. I have much experience with business process modelling in general.	3.19			
2. I have much experience with process mining in general.	2.63			
Model	**A**	**B**	**C**	**D**
Usefulness				
3. The information presented in this model is useful for my daily job.	1.94	3.13	3.56	3.19
4. The model is suitable for gaining new insights about the business process.	1.56	3.31	3.81	3.50
5. The model contains detailed information about the business process.	1.38	2.75	4.13	4.69
6. The model helps forming an understanding of the business process in general.	2.31	3.88	3.56	2.25
Average	**1.80**	**3.27**	**3.77**	**3.41**
Ease of Use				
7. The model is understandable when taking a first look at it.	4.00	3.56	2.88	1.56
8. It is easy to learn understanding this model.	4.06	3.50	3.06	1.75
9. It is easy to explain this model to other persons inside the organization.	3.75	3.69	3.06	1.50
10. Someone without experience in process mining is able to understand this model.	3.50	3.44	2.56	1.50
11. I will use the information obtained from this model in my daily job.	1.88	3.31	3.63	2.63
12. This model helps me achieve my purpose inside the organization.	1.69	3.31	3.81	2.94
Average	**3.15**	**3.47**	**3.17**	**1.98**

els all score above 3, with model C scoring the highest. The combination of abstracting specific behavior, whilst still providing enough detail, showed to be useful for the stakeholders.

As for the ease of use, the opinion of the stakeholders on models A and D has entirely changed. The fact that model D provides more information about the overall process, can make it challenging to understand for the stakeholders, which may result in a low score for ease of use. Models B and C still score consistently above 3, indicating that stakeholders also perceive these models as being (relatively) comprehensible and easy to use.

Using our understanding of the perceived usefulness, the perceived ease of use, and the quantitative metrics of every process model, an alignment of stakeholders and process model complexity can be made. The way in which the measurements of this alignment are used is situation dependent and one may favor one construct more than the other. Factors such as the type of stakeholder and the organizational context influence the preferred prioritization of each measurement. By excluding model C (e.g., because of its low fitness) and having discussions about the score on perceived usefulness and ease of use, authors of the present work arrived at the alignment made in Table 7. Our methodology allows to choose an approach in evaluating the constructs, depending on the context in which they apply our methodology. Defining a decision-making method (e.g., a ranking mechanism or any, more formal, form of multi-criteria decision-making method) is something that can be introduced in a future extension of the methodology.

Table 7: An example of stakeholders with preferable abstraction levels.

Stakeholder	**Abstraction level**	**Stakeholder**	**Abstraction level**
Operational board	B	**Exception manager**	D
CFO	B	**IT expert**	B
Planner	B	**Regulations expert**	D
Driver	B	**Customer relations**	B

Notice that we involved experts that reasoned from the perspectives of stakeholders, instead of a (group of) representative stakeholder(s). In principle, our proposed methodology could be applied to both experts and stakeholders. We decided to involve experts, as they were familiar with the semantics of process models and the concerned case study. Yet, not all experts were experienced with the stakeholder roles. However, we justify this choice by (1) using more than one expert reasoning on each stakeholder, (2) providing the experts a functional description of the stakeholders (Table 2), and (3) restating that our purpose concerns demonstrating a design artifact as part of a design science cycle instead of exhaustively evaluating a case study.

4 RELATED WORK

Events can be recorded at a (very) granular level, and if not dealt with appropriately (e.g., through process mining discovery algorithms) this can result in incomprehensible process models. What an "appropriate" level of detail is needed is debatable. Generated process models should be understandable (Van Cruchten and Weigand, 2018a). Mostly it is assumed that event data are of the same and bear an appropriate level of granularity. However, in reality, granular event logs often produce either spaghetti or lasagne process models (Van der Aalst, 2016). To understand

SCIENCE AND TECHNOLOGY PUBLICATIONS

these complex and (semi-)unstructured models there are many approaches proposed in literature. To mention some, there are pre-processing techniques that allow an appropriate level of granularity as for example identified by (Van Zelst et al., 2021) and approaches that use the complex process model and choose the level of details based on who is expected to benefit from the process model. The latter one can be exhibited in particular views for, e.g., customers (Bernard and Andritsos, 2018) or healthcare providers (Mans et al., 2012).

There exist many algorithms that focus on event log abstraction within the process mining discipline. For example, (De Medeiros et al., 2007) proposed a clustering algorithm for reducing the level of abstraction in process models. This algorithm iteratively makes clusters based on the event log, until the obtained process models do not over-generalize certain activities of the event log. (Becker and Intoyoad, 2017) explores how a k-medoids algorithm can be used to cluster heterogeneous datasets. They check the characteristics of the obtained models for different levels (of k). Their results indicate that it is useful to evaluate the resulting process models based on a specific purpose. However, a common value of k that is appropriate for several stakeholder purposes is not found. (Dos Santos Garcia et al., 2019) attempts to address this problem but also calls for future research on methods to determine the main processes for particular purposes (e.g., stakeholders). (Baier et al., 2014) uses an abstraction approach based on external domain knowledge. They stressed the importance of making process models understandable for business users, by working with an appropriate level of abstraction (Van der Aalst and Gunther, 2007; Günther and Van Der Aalst, 2007). (Fazzinga et al., 2018a) describe how event logs with low-level events that seem to have no reference to high-level activities, can be transformed to the preferred abstraction level of an analyst. Although their method helps with adjusting the event log such that it is more suitable for creating a process model, it does not specifically address how different abstraction levels influence the appropriateness of the process model for a given stakeholder. Other work from (Fazzinga et al., 2018c) proposes a framework that induces process models to describe the process at an activity level, to better suit the needs of process model analysts. Although this framework actively seeks to shift a process model towards an abstraction level that is more appropriate for model analysts, it does not elaborate on how distinct stakeholders require different levels of abstraction. In our study, we take explicitly into account the perspective of different stakeholders when assessing process models. In line with this, our work also relates to a client-server-based application proposed by (Yazdi et al., 2021) to gradually abstract fine-grained event logs to higher levels without losing essential information, thereby enabling the domain experts to use the appropriate process model for further analysis.

Our work also taps into a recent discussion on the use of agent-based modeling in combination with (data-driven) process mining techniques. The decision logic of the case study considered in the present paper uses agent-based modeling (Bemthuis et al., 2020), which acts as a natural recourse when incorporating human-interpretation capabilities or when interacting with humans. In our work, we give a qualitative and quantitative assessment of the resulting event logs which are originated from complex agent interactions. Thereby, we also put forward the call for research on addressing how agent-based models could affect the quality of process mined models (Bemthuis et al., 2019).

5 CONCLUSIONS AND FUTURE WORK

The relevance and comprehensibility of a process model for a particular stakeholder are greatly influenced by the level of abstraction of the process model. Yet, providing a suitable process model for a designated stakeholder can be challenging. Current literature on stakeholder analysis and process model abstraction in the process mining discipline acknowledges the need for aligning a stakeholders' purposes to relevant process models. Therefore, we proposed a methodology to align stakeholder needs with process model complexity. This methodology consists of six phases that each contribute to aligning the complexity of generated process models to the needs of particular stakeholders. A logistics case study, in which we used the TAM as a qualitative measurement, demonstrates the usefulness of our methodology for properly abstracting process models while taking into account the needs of stakeholders.

Some limitations and potential improvements in this study are as follows. The first is related to our case study, which involves only a few complex traces and the resulting process models were not exhaustive. Although we enriched the event logs by adding an additional attribute and considering multiple stakeholders, further work relies on more realistic and complex case studies to validate our methodology. Case studies in diverse domains and with representative groups of experts/stakeholders are desired to validate the presented methodology.

Second, the use of both quantitative and qualitative metrics to assess the quality of a process model from the perspective of different stakeholders is only explored to a limited extent. Future work could focus on how the interaction between a stakeholder/expert and mined process models will occur more specifically. Within the logistics domain, a suggestion could be to focus on process mining use cases involving the Open Trip Model (OTM). A conceptual mapping already outlined that OTM provides a promising way to unify storage, integration, interoperability, and querying of logistics event data (Piest et al., 2021).

Third, the application of TAM to reason on process model abstractions requires a deeper investigation of the role and expertise of experts/stakeholders. Further improvement lies in exploring possible ways to incorporate not only domain knowledge but also technical knowledge of the assessors. Our case study involved experts that reasoned on behalf of stakeholders, while our methodology may also be applicable to actual stakeholders in the field. The notion that the same identical process may have different representations for different stakeholders is an interesting outcome. We call for further work on the use of abstraction in process mining not only based on topology but also on content and the operational relevance of individual process steps. Hence, the graphical layout of a model should be separated from, or at least complementary to, its (knowledge) content. Recent developments on domain-knowledge-utilizing process discovery algorithms underscores this need (Schuster et al., 2022).

Fourth, it may also be fruitful to investigate how differentiation by means of abstraction levels (tailored for specific stakeholders) can be applied to other parts of a process mining project. For example, during the data processing stage or when evaluating a process mining project. Algorithms and methodologies could explicitly take into account multi-stakeholder views.

ACKNOWLEDGEMENTS

This work is part of the project DataRel (grant 628.009.015) which is (partly) financed by the Dutch Research Council (NWO). The authors would like to thank all project partners as well as Syeda Sohail for their support. The authors also thank the anonymous reviewers for their constructive feedback.

REFERENCES

Anthony, R. N. (1965). *Planning and control systems: a framework for analysis*. Division of Research, Graduate School of Business Administration, Harvard.

Baier, T., Mendling, J., and Weske, M. (2014). Bridging abstraction layers in process mining. *Information Systems*, 46:123–139.

Becker, T. and Intoyoad, W. (2017). Context aware process mining in logistics. *Procedia CIRP*, 63:557–562.

Bemthuis, R., Mes, M., Iacob, M., and Havinga, P. (2021a). Data underlying the paper: Using agent-based simulation for emergent behavior detection in cyber-physical systems. DOI: 10.4121/14743263. 4TU.ResearchData.

Bemthuis, R., Mes, M., Iacob, M.-E., and Havinga, P. (2020). Using agent-based simulation for emergent behavior detection in cyber-physical systems. In *2020 Winter Simulation Conference (WSC)*, pages 230–241. IEEE.

Bemthuis, R., van Slooten, N., Arachchige, J. J., and Piest, J. P. S. (2021b). A classification of process mining bottleneck analysis techniques for operational support. In *Proceedings of the 18th International Conference on e-Business*, pages 127–135. SCITEPRESS.

Bemthuis, R. H., Koot, M., Mes, M. R. K., Bukhsh, F. A., Iacob, M.-E., and Meratnia, N. (2019). An agent-based process mining architecture for emergent behavior analysis. In *2019 IEEE 23rd International Enterprise Distributed Object Computing Workshop (EDOCW)*, pages 54–64. IEEE.

Bernard, G. and Andritsos, P. (2018). CJM-ab: Abstracting customer journey maps using process mining. In *International Conference on Advanced Information Systems Engineering*, pages 49–56. Springer.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pages 319–340.

De Medeiros, A. K. A., Guzzo, A., Greco, G., Van der Aalst, W. M. P., Weijters, A. J. M. M., Van Dongen, B. F., and Sacca, D. (2007). Process mining based on clustering: A quest for precision. In *International conference on business process management*, pages 17–29. Springer.

Diba, K., Batoulis, K., Weidlich, M., and Weske, M. (2020). Extraction, correlation, and abstraction of event data for process mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3):e1346.

Dos Santos Garcia, C., Meincheim, A., Junior, E. R. F., Dallagassa, M. R., Sato, D. M. V., Carvalho, D. R., Santos, E. A. P., and Scalabrin, E. E. (2019). Process mining techniques and applications–a systematic mapping study. *Expert Systems with Applications*, 133:260–295.

Fazzinga, B., Flesca, S., Furfaro, F., Masciari, E., and Pontieri, L. (2018a). Efficiently interpreting traces of low level events in business process logs. *Information Systems*, 73:1–24.

Fazzinga, B., Flesca, S., Furfaro, F., and Pontieri, L. (2018b). Online and offline classification of traces of event logs on the basis of security risks. *Journal of Intelligent Information Systems*, 50(1):195–230.

Fazzinga, B., Flesca, S., Furfaro, F., and Pontieri, L. (2018c). Process discovery from low-level event logs. In *International Conference on Advanced Information Systems Engineering*, pages 257–273. Springer.

Flodén, J. and Woxenius, J. (2021). A stakeholder analysis of actors and networks for land transport of dangerous goods. *Research in Transportation Business & Management*, page 100629.

Gibson, K. (2000). The moral basis of stakeholder theory. *Journal of business ethics*, pages 245–257.

Graafmans, T., Turetken, O., Poppelaars, H., and Fahland, D. (2021). Process mining for six sigma. *Business & Information Systems Engineering*, 63(3):277–300.

Greenley, G. E. and Foxall, G. R. (1997). Multiple stakeholder orientation in uk companies and the implications for company performance. *Journal of Management Studies*, 34(2):259–284.

Gruhn, V. and Laue, R. (2007). Approaches for business process model complexity metrics. In *Technologies for business information systems*, pages 13–24. Springer.

Günther, C. W. and Van Der Aalst, W. M. P. (2007). Fuzzy mining–adaptive process simplification based on multi-perspective metrics. In *International conference on business process management*, pages 328–343. Springer.

Iacob, M.-E., Charismadiptya, G., van Sinderen, M., and Piest, J. P. S. (2019). An architecture for situation-aware smart logistics. In *2019 IEEE 23rd International Enterprise Distributed Object Computing Workshop (EDOCW)*, pages 108–117. IEEE.

Iacob, M. E., Monteban, J., Van Sinderen, M., Hegeman, E., and Bitaraf, K. (2018). Measuring enterprise architecture complexity. In *2018 IEEE 22nd International Enterprise Distributed Object Computing Workshop (EDOCW)*, pages 115–124. IEEE.

Kumar, M. V. M., Thomas, L., and Annappa, B. (2017). Distilling lasagna from spaghetti processes. In *Proceedings of the 2017 International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence*, pages 157–161.

Leemans, S. J. J., Goel, K., and van Zelst, S. J. (2020). Using multi-level information in hierarchical process mining: Balancing behavioural quality and model complexity. In *2020 2nd International Conference on Process Mining (ICPM)*, pages 137–144. IEEE.

Mans, R. S., Van der Aalst, W. M. P., Vanwersch, R. J. B., and Moleman, A. J. (2012). Process mining in healthcare: Data challenges when answering frequently posed questions. In *Process Support and Knowledge Representation in Health Care*, pages 140–153. Springer.

Muketha, G. M., Abd Ghani, A. A., Selamat, M. H., and Atan, R. (2010). A survey of business processes complexity metrics. *Information Technology Journal*, 9(7):1336–1344.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77.

Piest, J. P. S., Cutinha, J. A., Bemthuis, R. H., and Bukhsh, F. A. (2021). Evaluating the use of the open trip model for process mining: An informal conceptual mapping study in logistics. In *Proceedings of the 23rd International Conference on Enterprise Information Systems*, pages 290–296. SCITEPRESS.

Post, J. E., Preston, L. E., and Sachs, S. (2002). Managing the extended enterprise: The new stakeholder view. *California management review*, 45(1):6–28.

Schuster, D., van Zelst, S. J., and van der Aalst, W. M. P. (2022). Utilizing domain knowledge in data-driven process discovery: A literature review. *Computers in Industry*, 137:103612.

Tolentino-Zondervan, F., Bogers, E., and van de Sande, L. (2021). A Managerial and Behavioral Approach in Aligning Stakeholder Goals in Sustainable Last Mile Logistics: A Case Study in the Netherlands. *Sustainability*, 13(8).

Van Cruchten, R. M. E. R. and Weigand, H. H. (2018a). Process mining in logistics: The need for rule-based data abstraction. In *2018 12th International Conference on Research Challenges in Information Science (RCIS)*, pages 1–9. IEEE.

Van Cruchten, R. R. and Weigand, H. H. (2018b). Process mining in logistics: The need for rule-based data abstraction. In *2018 12th International Conference on Research Challenges in Information Science (RCIS)*, pages 1–9. IEEE.

Van der Aalst, W. M. P. (2016). *Process mining - Data science in action*. Springer.

Van der Aalst, W. M. P., De Medeiros, A. K. A., and Weijters, A. J. M. M. (2006). Process equivalence: Comparing two process models based on observed behavior. In *International conference on business process management*, pages 129–144. Springer.

Van der Aalst, W. M. P. and Gunther, C. W. (2007). Finding structure in unstructured processes: The case for process mining. In *Seventh International Conference on Application of Concurrency to System Design (ACSD 2007)*, pages 3–12. IEEE.

Van Eck, M. L., Lu, X., Leemans, S. J. J., and Van der Aalst, W. M. P. (2015). PM^2: A process mining project methodology. In *International Conference on Advanced Information Systems Engineering*, pages 297–313. Springer.

Van Zelst, S. J., Mannhardt, F., de Leoni, M., and Koschmider, A. (2021). Event abstraction in process mining: literature review and taxonomy. *Granular Computing*, 6(3):719–736.

Wirth, R. and Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*.

Wynn, M. T., Poppe, E., Xu, J., ter Hofstede, A. H., Brown, R., Pini, A., and Van der Aalst, W. M. P. (2017). Processprofiler3d: A visualisation framework for log-based process performance comparison. *Decision Support Systems*, 100:93–108.

Yazdi, M. A., Ghalatia, P. F., and Heinrichs, B. (2021). Event log abstraction in client-server applications. In *Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - Volume 1: KDIR,*, pages 27–36. SCITEPRESS.

Zakarija, I., Škopljanac-Mačina, F., and Blašković, B. (2015). Discovering process model from incomplete log using process mining. In *2015 57th International Symposium ELMAR*, pages 117–120. IEEE.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS