Finding the Appropriate Level of Abstraction for Process Mining in Logistics

Dennis Maneschijn
University of Twente
PO Box 217, 7500 AE Enschede
the Netherlands

d.g.j.c.maneschijn@student.utwente.nl

ABSTRACT

Process mining is a field within data science that focuses on the execution of business processes, as observed from real-life events. The mined process models often contain too much detail and, therefore, become hard to understand for various stakeholders. To solve this, abstraction could be applied. Abstraction is about simplifying a process model, to make it more comprehensible. It is often hard to find the right balance between having a model that gives actionable information, but that is not too detailed. Moreover, what level of abstraction is appropriate depends on what stakeholder is looking at the process model. This research defines multiple levels of abstraction and generates process models at each of these levels. Using quantitative measurements and an expert analysis we will reason on the quality of each process model for various stakeholders. We found that there exists a level of abstraction that is deemed appropriate for all stakeholders by both the quantitative analysis and the expert analysis. This research contributes to the existing research on abstraction in process mining, by explicitly defining a set of abstraction levels for the fuzzy miner. Moreover, we present how quantitative measures in combination with an expert analysis can be used to reason on the quality of a process model considering the needs of various stakeholders. This reasoning is used to define the most suitable abstraction level for every stakeholder.

Keywords

Process mining, abstraction, process discovery, process models, logistics, stakeholders

1. INTRODUCTION

Organizations aim to gain insights into their business processes and improve them wherever possible. Having efficient business processes is important for an organization, since it will help improve business results and remain competitive [9]. However, the existing business process models are often not in line with the actual process execution within the organization [28]. This makes it challenging for an organization, to understand where its process needs to be improved.

Process mining provides a solution for this problem [2]. Process mining is a relatively new field in data science that is becoming more and more popular [1]. Due to the digitalization of organizations a lot of information, so-called *event data*, has

become available. This event data is stored in *event logs*. Process mining uses these event logs to discover the way a business process is conducted in reality. Moreover, process mining can be used for conformance checking and in finding characteristics of the process, such as bottlenecks [2]. Process mining often shows that there exists a discrepancy between how the process is ought to be executed and how it is executed in practice [29]. Therefore, it can be valuable for organizations to apply process mining, since it gives them insights on many different aspects of their business process, that were not known before.

Logistical processes are known for being dynamic, heterogeneous, and human-centred [20]. This makes it important to use process mining, to get to know the insights of these processes and see where there is room for improvement [7, 22, 32]. Therefore, it is surprising that the literature on process mining in logistics still lacks behind many other industries [14]. This leaves plenty of room for further research on the application of process mining in logistics.

The complexity of logistical processes increases the likelihood of so-called spaghetti models [21]. To encounter this, abstraction can be applied to make the model less spaghetti-like and more comprehensible. Abstraction is about the level of granularity of the process model. A model can have too high abstraction (*underfitted*), or an abstraction that is too low (*overfitted*) [4]. It is often hard to find the right balance between a model that is not too detailed, but that does provide enough information to the reviewer [4, 23]. Moreover, the right level of abstraction also depends on which stakeholder is using the model, since stakeholders have different needs and purposes [13, 30, 31]. Multiple papers address the need for abstraction in process mining for logistics or they propose methods for abstraction in general, but none of these try to seek an abstraction level that serves the purpose of a stakeholder in the organization [6, 7, 10, 18, 24, 25].

The aforementioned lack of related work shows that there is plenty of room for new research in this area. The goal of our research is to analyse what level of abstraction in process models is most suitable for the different purposes of the involved stakeholders. In the end, we try to discover a generic level of abstraction that is suitable for all stakeholders inside a logistical organization. In order to reach our goal, we have defined three research questions:

- **RQ1:** What abstraction method is most suitable for process mining in logistics and what levels of abstraction can we define for this method?
- **RQ2:** What relevant stakeholders can be defined in a logistical organization and what purposes do they have?
- **RQ3:** How can we identify whether a specific model provides enough information to serve the purposes mentioned in **RQ2**?

When we know the most suitable abstraction method with its corresponding abstraction levels (**RQ1**), the purposes of stakeholders (**RQ2**), and the ways to check whether a model is sufficient enough (**RQ3**), we can analyse what abstraction level generates an appropriate model when process mining is applied to our dataset.

The remainder of this paper is structured as follows. Section 2 discusses some background information that is relevant for this research. The methodology of this research is discussed in section 3. Section 4 provides the findings of the literature research. In section 5, the results of the experiment are reported. Section 6 reasons on the quality of this research and discusses certain threats to validity. Finally, section 7 concludes and reflects on the goal of this research.

2. BACKGROUND
In order to perform this research, it is important to know some fundamental concepts, discussed in this paper. Understanding these concepts helps to interpret the techniques used for the experiment and to identify the quality of the model.

2.1 Abstraction in Process Mining
As described in [12], a process mining project typically includes six steps. A schematic overview of these steps can be found in Figure 1. Many of these steps include ways to perform abstraction. The goal of this abstraction is to, in the end, simplify the generated process model. This is often necessary since real-life event logs tend to generate spaghetti models. Since abstraction helps simplifying the model, the spaghetti model will be converted into a so-called lasagna model that is comprehensible for the person looking at it [24].
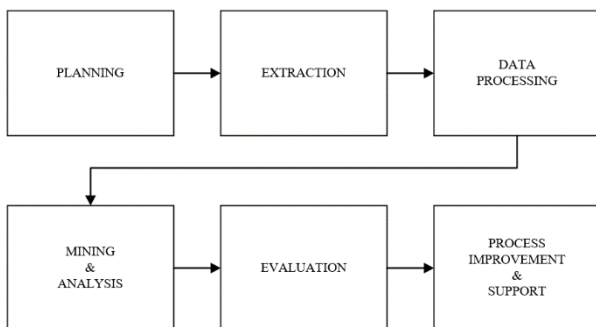

Figure 1. PM² methodology [12]**.**

First of all, in the extraction of event data, abstraction can be performed. One can for example decide to omit certain activities from the event log, to make it more simplistic. Also, the number of metadata captured can be reduced. Van Cruchten and Weigand [10] describe how data extraction by information systems can play an important role in simplifying the process model.

Another popular place to perform abstraction is the data processing phase. Once the event logs are filled with event data, we can apply many algorithms that help with simplifying the event logs. For example, De Medeiros [25], uses a clustering algorithm for reducing the level of abstraction in process models. They iteratively keep clustering the event log until all process models do not over-generalize some certain actions of the event log. Moreover, Becker and Intoyoad [7], explore how the k-medoids algorithm can be used to cluster heterogeneous logistical datasets. They show that it is possible to use this algorithm and that it can be useful to evaluate the resulting models based on a specific purpose. They also check the characteristics of the models for different levels of k; however, they do not try to find a common value of k that is the most

appropriate for several stakeholder purposes. They suggest that finding a common value of k is something that could be done in future research.

The final phase in which abstraction can be performed is the mining & analysis phase. Baier et al [6] propose a method for using abstraction in process models, by using external domain knowledge. They also state the importance of making process models understandable for business users, by working with the right level of abstraction. Other papers that describe the use of abstraction in process discovery, are the papers of Günther and van der Aalst [4, 18]. Using the fuzzy miner algorithm in ProM, they can simplify process models using certain thresholds inside the algorithm.

This research focuses on the application of abstraction in the mining & analysis phase of process mining. The main reason for this, is that we want to make sure our research is as relevant as possible. By reviewing abstraction in the last part of the process, it does not matter what abstraction has been performed in the beginning, since we review how any event log can be used to make a model with different levels of abstraction. We want to give the end user the possibility to look at the different levels and decide what level suits them the best. To delineate this research, we will define our definition of abstraction that we will be using in this research. In [4] abstraction is defined as omitting lower-level information, that is insignificant in the chosen context, from the visualization. For example, small roads and bicycle paths are not shown when looking at a city road map, since it would make the map cluttered and inconvenient to use. Using this explanation, we have defined the following definition of abstraction in process mining:

"*Simplifying models by removing edges, clustering nodes, and removing nodes to make the process model more comprehensible for the person looking at it.*"

It is often challenging to find the right balance between a comprehensible model and a model that provides enough information to be useful [23]. This research tries to find this balance by considering multiple stakeholder views.

2.2 Dataset
The data used in the experiment is the dataset used in the work of Bemthuis et al. [8]. The case study concerns a simulation of a logistics process and stored important event data in event logs. We will shortly explain the fundamentals of their simulation study and resulting event logs. The simulation contains three types of vehicles: Unmanned aerial vehicles (UAV), human-driven forklifts (HDF), and automated guided vehicles (AGV). Every type of vehicle has its own travel time and amount of product decay. Every product that enters the pipeline has a decay level of 100. Over time this level will decrease. We can interpret this decay level as a quality measure, the higher the better. After the products have been transported to the start of the production line, they will follow a certain production process. This process is ought similar for every product and therefore not included in the event log. After the production process, the products are picked up again by one of the three vehicle types and transported to the end destination. We used an event log that included a so-called 'warmup filter', this means that the simulation has run for a period before it started logging events. Out of the 540 simulation runs, we selected a run from the scenario that had the lowest average decay, to ensure relevant event logs were used. A more detailed description of the process can be found in [8].

2.3 Process Miner
In process mining, we use process miners to generate process models. A miner is an algorithm that generates a model from a given event log. The available miners all use different techniques

and produce different kinds of models. Therefore, which miner is most suitable, highly depends on the context in which the process mining is applied and the quality of the event log. Two well-known process discovery miners are the heuristics miner and the fuzzy miner. The heuristics miner is one of the traditional process miners that can deal well with noise in logs. It was designed to solve existing problems in the alpha-miner and make it more suitable for real-life event logs [33]. The fuzzy miner is closely related to the heuristics miner. However, the fuzzy miner uses a unique combination of *significance* and *correlation* thresholds to easily simplify the resulting model [4, 18]. This makes the fuzzy miner suitable for highly complex and unstructured event logs. As mentioned in section 1 of this research, logistical processes are characterized for being highly dynamic and heterogeneous. Therefore, we have decided to use the fuzzy miner in this research. Moreover, the relatively easy simplification of process models allows us to define certain abstraction levels for which we can generate different models. On top of this, the fuzzy miner is suitable for mining reality, since it allows for edges that are not necessarily expected. These edges can be relevant for certain stakeholders.

2.4 Quality Measurements

In process mining, when we want to reason on the quality of process models, we can either use *quantitative* measures or *qualitative* measures. One popular quantitative measure in process mining is *fitness*. Fitness measures how well the behaviour inside an event log is displayed in the process model [3]. A high fitness means that the model is properly representing the event log, while a low fitness score means that the model is not properly showing what behaviour is observed. Fitness is useful for getting an initial idea of the quality of the model. If we know that the model is not properly displaying reality, we will also be less interested in some other details of the models, since it is less relevant. Apart from the fitness, several statistics of the process model are used to give the reader a general understanding of the complexity of the process models. The first statistic is the *level of detail* inside the process model. This is a percentage that displays how many nodes are preserved in the model. The other statistics include the number of nodes, edges, and clusters inside the model.

On the other hand, we have qualitative measures. A well-known qualitative analysis methodology, is the Technology Acceptance Model (TAM) [11]. TAM defines a set of questions that can be asked to users, to measure their *perceived usefulness* and *perceived ease of use* of a new technology. The first one is the degree to which a person thinks the system enhances his job performance. The second one is about the degree to which a person thinks, using the system will be free of effort. Users can give scores from 1 to 5 to these questions, that can later be analysed and drawn conclusions from. Originally, TAM was designed for measuring the usefulness of computer systems. We will adapt the questions to make them more suitable for process mining and the reviewing of process models. We will perform an expert analysis in which people will take the stance of different stakeholders and answer our questions. In the end, we can use this expert analysis to draw conclusions on the quality of the model.

3. METHODOLOGY

This section will shed some light on how the research was performed and what steps were taken in order to reach the main goal of this paper. First of all, a literature research about abstraction in process mining and how this abstraction changes the outcome of process models was done. Also, papers on process mining in general were analysed to create a solid

foundation of knowledge on process mining. To consolidate this foundation even further, an online Coursera course on process mining was followed. Google Scholar and Scopus were used to find relevant papers. Moreover, the systematic mapping study of Garcia, C. dos S. et al. [14] was used to find papers on specific topics.

3.1 CRISP-DM

To make sure a structured process was followed during the rest of the research, we used CRISP-DM [34] as a standard. CRISP-DM is a well-known framework for carrying out data mining projects. A schematic overview of CRISP-DM including the corresponding sections of this research can be found in Figure 2.
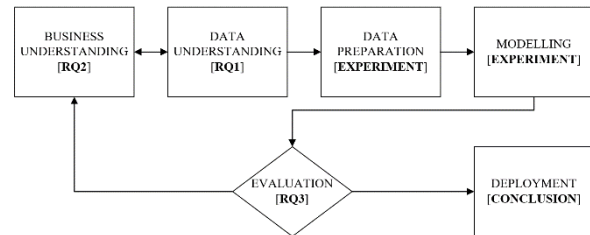

Figure 2. CRISP-DM methodology overview [34]**.**

3.1.1 Business & data understanding

We used RQ1 and RQ2 to build the business and data understandings. This started with looking through the event logs and reading the paper of Bemthuis et al. [8], to get to know the insights of the simulation. Afterwards, we also constructed some process models using the Disco software, to identify the general patterns in the process models. Moreover, a brief literature review was done to choose a proper miner for which we could eventually define abstraction levels. Eventually, this data understanding lead to the selection of the fuzzy miner and a definition of several abstraction levels. Simultaneously we also worked on the business understanding, which allowed us to create a list of stakeholders, each with a different purpose inside the organization. To do this multiple stakeholder analyses in the logistics sector were read to form a basic understanding of the important stakeholders. In the end, we used the business and data understandings to create a matching between the abstraction levels and the stakeholders. This matching is used as a hypothesis in the deployment phase.

3.1.2 Data preparation

The data preparation started with enriching the data. To achieve this, two attributes inside the event log were used. These attributes are the decay level *(DL)* of a product, and the vehicle that performs a certain task. In order to make these quality levels visible inside the model, we classify the decay levels into four different categories. These categories are based on the mean and the standard deviation of the decay level inside a simulation run. The categories can be found in table 1. We have decided to not use generic quality levels, since we wanted to make sure that the events are evenly distributed over the different quality levels. After enriching the data, we also filtered the event log, to remove incomplete traces. Filtering was done using the heuristics filter in ProM.

Table 1. Quality levels used for data enrichment

Quality level	Partition of event log
Good	$DL >= \mu + \sigma$
Sufficient	$DL >= \mu$ and $DL < \mu + \sigma$
Insufficient	$DL < \mu$ and $DL > \mu - \sigma$

Poor	$DL <= \mu - \sigma$

3.1.3 Modelling

According to CRISP-DM, the modelling phase usually includes choosing an appropriate modelling technique (miner). This research, however, defines abstraction levels that depend on the chosen miner. Therefore, we decided to choose a miner at the beginning of the research. During the mining phase we will generate multiple models that each represent a different level of abstraction. All process models were mined using ProM.

3.1.4 Evaluation & deployment

In this research we will evaluate whether a certain abstraction level provides the right information for a particular stakeholder. To do this, we will both perform quantitative and qualitative measurements. The results of these measurements will give us the possibility to reason on the quality of each process model. The goal of the measurement is to, in the end, define what level of abstraction is best for every stakeholder. Eventually, the results of the evaluation will be used to reflect on the hypothesis, created in RQ2. Also, we expect to discover some general patterns on the preferred level of abstraction. These patterns will help us provide a conclusion to this research and serve as a deployment for future research.

4. FINDINGS

This section will mention the relevant findings for each research question. These findings are later on used for the experiment.

4.1 Establishing abstraction levels (RQ1)

As mentioned in section 2.3, we will use the fuzzy miner for generating our process models. The next step is to define suitable abstraction levels at which we will generate the models. We mentioned that the fuzzy miner uses thresholds that determine whether activities are displayed in the model, based on their significance and correlation.

4.1.1 Significance & correlation

According to [4], *significance* is about the relative importance of behaviour. This relative importance can both be measured for the event classes (activities) and the relation between them (edges). In other words: "significance specifies the level of interests we have in events, or their occurrence after one another". Generally, activities that occur a lot in the event log are considered more significant. *Correlation* is only about the precedence relation of two events. The more closely related two events following each other are the higher their correlation. To measure correlation, we consider the amount of data attributes that are associated to two events that follow each other [4].

4.1.2 Fuzzy miner thresholds

Using the understanding of significance and correlation, we can explain the different thresholds used by the fuzzy miner for model simplification.

The first step of simplifying is about *conflict resolution*. Nodes conflict when edges exist from both nodes A to B and from nodes B to A. Our dataset only contains sequential traces without concurrency and loops. Therefore, conflict resolution and its corresponding thresholds are outside the scope of this research.

After the conflict resolution, we continue with filtering all other irrelevant edges. First, the *utility* of every edge is determined using the **utility ratio.** After that, we remove all edges that have a *normalized utility* below a certain **edge cutoff threshold.**

At last, we perform simplification on the nodes. All nodes that have *unary significance* below the **node cutoff threshold** will be placed in clusters. All clusters that are completely *isolated* from

the model (no edges from or to it) and the clusters that are *singular* (contain only one node), will be removed from the model. A more extensive explanation can be found in [4].

The following table summarizes the thresholds found:

Table 2. Relevant thresholds in 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

In general, we can see that the lower our thresholds become, the less abstract our model will be. The same holds the other way around. The utility ratio, however, does not directly influence the abstraction of the model since it only focusses on a mixture of significance and correlation. Therefore, we will keep its value constant, to make sure it will not bias our results. Moreover, due to the enrichment of the data, there are not many activities with a high significance. Therefore, we kept the NC at a relatively low level, to prevent the model from containing only one cluster. This results in the following abstraction levels:

Table 3. 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

To summarize, we will use the fuzzy miner for our experiment. We will use the generated process models, to reason on the appropriateness of the abstraction level used. In total, there are four abstraction levels (A, B, C, D). Abstraction level A is the most abstract and, therefore, contains the least details. Abstraction level D is the least abstract.

4.2 Identifying relevant stakeholders (RQ2)

Now that we have defined what miner we will use and how we will use different abstraction levels, it is time to identify a list of relevant stakeholders in logistics. For each stakeholder in the list, we will check the appropriateness of the process models. Actively managing stakeholders and addressing the needs of stakeholders is beneficial for an organization [17, 27]. As mentioned before in this research, process mining can also drastically improve business processes. However, the usefulness of process mined models depends on, e.g., whether the model is comprehensible for the stakeholder looking at it. Therefore, we will define a purpose for each stakeholder on the list and match this purpose to one of the four abstraction levels, defined in RQ1. This initial matching will constitute the hypothesis of our research. During the experiment, we will generate the process models and check whether our initial matching between abstraction levels and stakeholders was correct.

The first step in the process of creating a stakeholder list, is looking for relevant papers. There are quite some stakeholder analyses available for the logistics sector. One thing that these have in common, is that they both include primary and secondary

stakeholders [13, 19, 30, 31]. Primary stakeholders have a formal, or contractual relationship to the organization, while secondary stakeholders are not directly connected to the company [15]. Examples of external stakeholders are citizens and governmental regulators that are involved in a logistical process. Since these secondary stakeholders are not really bothered with analysing the process model, we will not put them directly into our stakeholder list. Instead, we will define primary stakeholders that represent the needs of secondary stakeholders.

Stakeholders have different goals and interests [13, 30, 31]. Due to these differences, different stakeholders are concerned with different kinds of information. It might be the case that one stakeholder wants to know a lot about the overall structure of the process, while another stakeholder is more interested in very specific events. For each stakeholder in our list, we will define a general purpose that is deemed relevant for him.

Monsalve et al. propose different models [26] that describe the abstraction of process models. One model they mention is Anthony's model [5]. Anthony defines three levels of abstraction inside an organization. The strategic level is the top layer that includes the general business goals and objectives. The tactical level is about obtaining and efficiently using resources. Finally, the operational level is about the efficient execution of specific tasks. We could see the strategical level as the most abstract layer and the operational layer as the most specific layer. Therefore, we can say that level A is on the strategic level, B and C are on the tactical level and D is on the operational level.

We will use the purposes of our stakeholders and the relation between Anthony's model and our abstraction levels, to create a hypothetical matching of stakeholders and abstraction levels.

Table 4. Stakeholders with abstraction levels based on Anthony's model

Stakeholder	Purpose	Abstraction level (hypothetical)
Operational board	Identify the overall workflow of the company.	A
CFO	Get to know the overall cost picture and identifying the specific causes of high costs.	A, C
Planner	Identify all the steps an order undergoes and where delays occur.	B
Chauffeurs	Find out what activities constitute to their specific task.	D
Exception manager	Spot exceptions and find out how they occurred.	D
IT expert	Find out what parts of the process require more extensive logging.	C
Regulations expert	Make sure all steps necessary for regulation measures are taken.	D
Customer relations	Ensure traceability and timeliness of the orders.	D

Our list contains a few stakeholders that are grouped because they share the same goals. First of all, the operational board mainly represents the CEO and COO. They are interested in how the entire process inside the organization is conducted and where there is room for improvement based on efficiency. Secondly, the chauffeurs are a group of both truck drivers and machinists, but also the people working inside a warehouse. They are both interested in their specific activity and how their activity relates to its neighbours. However, they seem to only benefit limitedly from knowing what the entire process looks like. Finally, the regulations expert and customer relations are representatives of secondary stakeholders. The regulations expert is concerned with governmental laws and certain restrictions given by local authorities. Customer relations on the other hand is interested in enhancing the customer journey, by ensuring traceability and timeliness of orders.

4.3 Defining quality measurements (RQ3)

In this third research question, we will focus on defining a method that allows us to check which of the generated process models is most suitable for a stakeholder. Since each process model belongs to a specific level of abstraction, we can use this method to also identify what abstraction level is most suitable for a particular stakeholder. This will eventually help us get to our final goal: defining to what extent there exists a common level of abstraction that is most suitable for all stakeholders inside a company.

In section 2.3 we highlighted how both quantitative and qualitative measures will be used. The quantitative measures will serve as a rough initial indication of the quality of the model. Especially the fitness is useful, since it directly measures how well the process model replays the event log. Also, the number of nodes, edges, and clusters makes it easy to quickly compare the complexity of the process models.

In order to perform the qualitative analysis an expert analysis in combination with a technology acceptance model (TAM) is used [11]. An expert analysis consists of a set of 'experts' that evaluate the process models, while reasoning from the perspective of a certain stakeholder. By mimicking a certain stakeholder, we know the pros and cons of the different process models. In the end, these pros and cons can be used to draw conclusions on the appropriateness of the models. In order to get sound results, we conduct the expert analysis in combination with TAM. As mentioned in section 2.4, 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 and has been adopted by many other articles [16, 35]. We will use these adaptations to create our own TAM.

We will first ask the respondents whether they are experienced in both business process management and process mining. This information is useful, since it allows us to know what model is suitable for inexperienced users. After this generic information, we want to know for each process model how useful it is, as perceived by the user. Important points for measuring the usefulness are:

- Effectiveness in daily job
- Possibility to gain new insights
- Possibility to identify specific instances
- Possibility to oversee the general process

The next step in TAM is to get to know more about the ease of use of each process model, as perceived by the user. Ease of use depends on the following aspects:

- Understandability of the model
- Ease of learning the model
- Explain ability of the model
- Experience needed for understanding the model

Finally, we want to identify the intent of the user. This can be measured using the following concepts:

- Intend to use the model

- Likelihood the model supports stakeholder purpose

Using these pillars, we have defined a set of questions that form our TAM. The resulting framework can be found in appendix A. Every stakeholder is reviewed by two experts. Each expert had to give a score from 1 to 5 for each process model related to the statement. The scores of two are displayed as 'a;b'. In total there are four of these 'a;b' ratings for every stakeholder for every statement. Each combination is for a different model. So, top left corresponds to level A, top right to level B, bottom left for model C, and bottom right for model D. The average scores for every model can be found in the column on the far right and the bottom row.

5. RESULTS

For the evaluation phase, we will first perform a quantitative analysis using the measurements defined in section 2.4. The fitness score is the most important to reason on the quality of the model. The other statistics help form a general understanding of the complexity of the model. The results of this quantitative analysis can be seen in table 5.

Table 5. Quantitative analysis

Model	Fitness	Detail	Nodes	Edges	Clusters (nodes)
Original	95.34%	100%	73	80	0
Level A	99.84%	52.44%	13	21	3 (8, 24, 28)
Level B	96.26%	69.17%	20	40	5 (5, 12, 24, 8, 3)
Level C	91.99%	88.71%	38	93	3 (4, 21, 2)
Level D	99.00%	100%	73	150+	0

The first thing that can be seen is that both the original model and model D include all possible events. However, their fitness scores differ wildly. The main reason for this is the difference in the number of edges present in the models. Since there are a lot more edges visible in model D, a lot more specific traces are possible according to the model. Although the model might be perceived as cluttered, it could provide interesting information for certain stakeholders. It is often not clear what happens inside the clusters. Therefore, the fact that no clusters are present in model D could improve its understandability. Whether this is also perceived by the stakeholders will become clear in the qualitative analysis. When we look at model C, we see that the fitness score is low. This is probably caused by the fact that the number of nodes in clusters is still quite low, and the number of edges has also been reduced drastically compared to model D. We will take the low fitness into account when drawing conclusions, since a low fitness is not desirable in a process model. Model B seems to strike a nice balance between showing a good number of nodes and edges, but also reduce a lot of irrelevant traces. Its fitness is not high, but it is acceptable. Moreover, the fact that model B contains multiple smaller clusters, could also imply a higher perceived usefulness. The reason for this, is that it can be unclear what happens inside large clusters. Finally, we will look at model A. Model A clearly provides the least detail, with only showing 13 out of the 73 activities. Almost all nodes are captured inside clusters, which causes the fitness to be high. The qualitative analysis will show whether the lack of detail in the model results in being less useful.

The results of the expert analysis in combination with the TAM can be seen in appendix A. In general, we see that model B and

C are preferred the most. Stakeholders perceive the models as both useful and easy to use. When looking at the scores of each statement, we see that models B and C either score the highest or score right in between models A and D. Also, when we look at the scores of all the individual stakeholders, we see that almost all stakeholders rank models B and C significantly higher than models A and D. One stakeholder that does not follow this rule is the exception manager. His score for model D is almost as high as the scores of models B and C. A good explanation for this is that model D allows for a lot of specific traces, due to its large number of edges, as mentioned in table 5. Moreover, the exception manager is deemed to have more experience with process mining than the other stakeholders. This might also explain why he is able to cope better with the complexity of model D. All the other stakeholders did not share the same opinion as the exception manager. Overall, the perceived usefulness of model A was fairly low, while its ease of use was deemed quite high. This suggests that the model does not provide enough information, but that it is comprehensible. Model D on the other hand, scores low on the perceived ease of use, while its perceived usefulness is ranked quite well. This suggests that again the right balance between being a comprehensible model while showing enough information was not found by the process model.

When combining the results of both the quantitative and qualitative results, we can draw some general conclusions. First of all, we found that the high level of detail found in model D, indeed resulted in being less comprehensible for most stakeholders. Moreover, the lack of detail in model A resulted in it being too abstract and not informative for the stakeholders. All stakeholders perceived this, and therefore ranked models B and C highest. Although model C is ranked a tiny bit higher than model B, we do need to consider the quantitative measure. It shows us that the fitness of model C is low, which means model C is not properly displaying reality. Taking all the aforementioned results into consideration allows us to create a final matching between all stakeholders and abstraction levels. This final matching can be seen in table 6.

Table 6. Stakeholders with most suitable abstraction levels

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

6. DISCUSSION

This discussion will reflect on the overall approach taken in this research. We will highlight some aspects that could have been performed in a different way, to ensure quality of our research would have increased.

One of the main threats to the validity of this research is the dataset used. The dataset was generated by a simulation model and contained few complex traces. Therefore, the resulting process models were straightforward and not really suitable for an extensive analysis. To make the event log more detailed, the data was enriched by making a difference in vehicle type and quality level. This data enrichment influenced the characteristics of the process. One could argue that, therefore, the original process was somewhat lost, and the quality of the dataset decreased. It is desirable to have authentic richness inside a dataset since it prevents from losing the original characteristics of the dataset.

Although we tried to make sure the abstraction levels we defined are as general as possible, it could be that the levels are quite specific for our dataset. Our dataset contains a lot of insignificant activities, which were mainly caused by the data enrichment. Therefore, the node cutoff is kept relatively low, to make sure the models would not consist of only one cluster. It could be the case that other datasets require abstraction levels with different threshold values. We discuss this further in the next section.

Another thing that could have been more refined, is the stakeholder list. The current list was established by reviewing multiple stakeholder analyses in the logistical sector. A better approach would have been to go to multiple organizations and identify their relevant stakeholders. Moreover, it would have been ideal to get a dataset from the same source as the stakeholders. This way, we certainly know that the process models will contain information that is related to the stakeholders. Our dataset did not really include information on the legal aspects; therefore, the reasoning of the regulation expert sometimes became a bit vague. Having a dataset from an organization with a regulation expert could have solved this.

Finally, we used an expert analysis for performing the TAM. It would have been better to conduct the TAM at an actual logistical organization. Currently, there is a risk that our experts did not properly reason from the stance of a particular stakeholder. Conducting the TAM at an actual logistical organization solves this and would have made the results less prone to errors. Also, more experts should have been approached to make the results more refined.

7. CONCLUSIONS AND FUTURE WORK

Process mined models often contain a lot of detail and are therefore hard to understand for various stakeholders. Abstraction can be used to solve this. Therefore, the goal of this research is to show whether there exists a common level of abstraction that is deemed appropriate for multiple stakeholders inside a logistics company. We found that there exists a level of abstraction that was found appropriate in both the quantitative and qualitative analyses performed. The process model corresponding to this abstraction level seemed to strike a right balance between being comprehensible and providing suitable information. The process models belonging to the other abstraction levels scored either low in the quantitative or the qualitative analysis. We also found that stakeholders tend to avoid process models that are easy to understand but provide no clear insights; and process models that provide a lot of detail but are hard to understand.

In general, we found that it is possible to have a single level of abstraction that is relevant for many different stakeholders. We also showed how conclusions like this can be drawn when considering both the quantitative and qualitative measurements. We did find one outlier from our general theory. There was a stakeholder who preferred the most detailed process model, because of the rich information it provides. This is something we expected, considering the purpose of the stakeholder.

The first research question showed that certain thresholds, used by the fuzzy miner, are suitable for defining multiple levels of abstraction in process mining. We also showed how the node cutoff should be weighed less, in order to make sure the process models would still be relevant. Finally, we saw how the edge cutoff plays a major role in simplifying a process model.

The second research question identified the most important stakeholders inside a logistical organization. It also used Anthony's model to create a hypothetical matching between stakeholders and abstraction levels. As shown in section 5, this hypothesis was not really correct, since we expected more stakeholders to prefer models A and D. However, the qualitative analysis showed us that most stakeholders found these models too abstract or too incomprehensible respectively. We saw that models B and C were perceived as the most suitable. They both correspond to the tactical level of Anthony's model.

Finally, the third research question showed how both a quantitative and qualitative analysis can be used to reason on the quality of process models for different stakeholders. We used existing quantitative measures available in ProM and designed our own technology acceptance model, to perform a qualitative analysis.

Future research should focus on performing this research on different event logs. Preferably, event logs that contain real-life events and have a lot of detail such as loops and concurrency. We are also interested to see whether the abstraction levels we have defined still hold for different event logs. Moreover, we think that it is important to create a stakeholder list that is related to the dataset. Another thing we think is interesting for future directions is the use of different process miners. We limited ourselves to the fuzzy miner and defined abstraction levels, specifically for the fuzzy miner. It will be interesting to see how different abstraction levels can be applied to different process miners. Finally, one other direction for future research is the application of abstraction levels in different steps of the process mining process as described in section 2. For example, how different abstraction levels can be used in the extraction and preparation of data. This can be combined with performing this research in different fields. We limited ourselves to the logistical sector, but it could be interesting to see how stakeholders in other industries have different opinions on the abstraction levels.

8. ACKNOWLEDGEMENTS

9. REFERENCES

1. Van Der Aalst, W. et al.: Process mining manifesto. In: Lecture Notes in Business Information Processing. (2012). https://doi.org/10.1007/978-3-642-28108-2_19.

2. van der Aalst, W.M.P.: Process Mining. Springer Berlin Heidelberg, Berlin, Heidelberg (2011).

https://doi.org/10.1007/978-3-642-19345-3.

3. Van Der Aalst, W.M.P. et al.: Process equivalence: Comparing two process models based on observed behavior. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). (2006). https://doi.org/10.1007/11841760_10.

4. Van Der Aalst, W.M.P., Günther, C.W.: Finding Structure in Unstructured Processes: The Case for Process Mining. In: Proceedings - 7th International Conference on Application of Concurrency to System Design, ACSD 2007. (2007). https://doi.org/10.1109/ACSD.2007.50.

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

6. Baier, T. et al.: Bridging abstraction layers in process mining. Inf. Syst. 46, (2014). https://doi.org/10.1016/j.is.2014.04.004.

7. Becker, T., Intoyoad, W.: Context Aware Process Mining in Logistics. In: Procedia CIRP. (2017). https://doi.org/10.1016/j.procir.2017.03.149.

8. Bemthuis, R. et al.: Using Agent-Based Simulation for Emergent Behavior Detection in Cyber-Physical Systems. In: Proceedings - Winter Simulation Conference. (2020). https://doi.org/10.1109/WSC48552.2020.9383956.

9. Carpinetti, L.C. r. et al.: Quality management and improvement: A framework and a business-process reference model. Bus. Process Manag. J. 9, 4, (2003). https://doi.org/10.1108/14637150310484553.

10. Van Cruchten, R.M.E.R., Weigand, H.H.: Process mining in logistics: The need for rule-based data abstraction. In: Proceedings - International Conference on Research Challenges in Information Science. (2018). https://doi.org/10.1109/RCIS.2018.8406653.

11. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. Manag. Inf. Syst. 13, 3, (1989). https://doi.org/10.2307/249008.

12. Van Eck, M.L. et al.: PM2: A process mining project methodology. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). (2015). https://doi.org/10.1007/978-3-319-19069-3_19.

13. Flodén, J., Woxenius, J.: A stakeholder analysis of actors and networks for land transport of dangerous goods. Res. Transp. Bus. Manag. (2021). https://doi.org/10.1016/j.rtbm.2021.100629.

14. Garcia, C. dos S. et al.: Process mining techniques and applications – A systematic mapping study, (2019). https://doi.org/10.1016/j.eswa.2019.05.003.

15. Gibson, K.: The moral basis of stakeholder theory, (2000). https://doi.org/10.1023/A:1006110106408.

16. Graafmans, T. et al.: Process Mining for Six Sigma: A Guideline and Tool Support. Bus. Inf. Syst. Eng. (2020). https://doi.org/10.1007/s12599-020-00649-w.

17. Greenley, G.E., Foxall, G.R.: Multiple stakeholder orientation in UK companies and the implications for company performance. J. Manag. Stud. 34, 2, (1997).

18. Günther, C.W., Van Der Aalst, W.M.P.: Fuzzy mining - Adaptive process simplification based on multi-perspective metrics. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). (2007). https://doi.org/10.1007/978-3-540-75183-0_24.

19. Iacob, M.E. et al.: An architecture for situation-aware smart logistics. In: Proceedings - IEEE International Enterprise Distributed Object Computing Workshop, EDOCW. (2019). https://doi.org/10.1109/EDOCW.2019.00030.

20. Intayoad, W., Becker, T.: Applying Process Mining in Manufacturing and Logistic for Large Transaction Data. In: Lecture Notes in Logistics. (2018). https://doi.org/10.1007/978-3-319-74225-0_51.

21. Jagadeesh Chandra Bose, R.P., Van Der Aalst, W.M.P.: Abstractions in process mining: A taxonomy of patterns. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). (2009). https://doi.org/10.1007/978-3-642-03848-8_12.

22. Kurganov, V. et al.: Process Mining as a Means of Improving the Reliability of Road Freight Transportations. In: Transportation Research Procedia. (2021). https://doi.org/10.1016/j.trpro.2021.02.076.

23. Leemans, S.J.J. et al.: Using multi-level information in hierarchical process mining: Balancing behavioural quality and model complexity. In: Proceedings - 2020 2nd International Conference on Process Mining, ICPM 2020. (2020). https://doi.org/10.1109/ICPM49681.2020.00029.

24. Manoj Kumar, M. V. et al.: Distilling lasagna from spaghetti processes. In: ACM International Conference Proceeding Series. (2017). https://doi.org/10.1145/3059336.3059362.

25. De Medeiros, A.K.A. et al.: Process mining based on clustering: A quest for precision. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). (2008). https://doi.org/10.1007/978-3-540-78238-4_4.

26. Monsalve, C. et al.: BPM and requirements elicitation at multiple levels of abstraction: A review. In: Proceedings of the IADIS International Conference Information Systems 2011, IS 2011. (2011).

27. Post, J.E. et al.: Managing the extended enterprise: The new stakeholder view, (2002). https://doi.org/10.2307/41166151.

28. Rozinat, A., van der Aalst, W.M.P.: Conformance checking of processes based on monitoring real behavior. Inf. Syst. 33, 1, (2008). https://doi.org/10.1016/j.is.2007.07.001.

29. Soffer, P. et al.: From event streams to process models and back: Challenges and opportunities. Inf. Syst. 81, (2019). https://doi.org/10.1016/j.is.2017.11.002.

30. Taylor, M. a P.: The City Logistics paradigm for urban freight transport. Infrastructure. 18, (2006).

31. Tolentino-Zondervan, F. et al.: A Managerial and Behavioral Approach in Aligning Stakeholder Goals in Sustainable Last Mile Logistics: A Case Study in the Netherlands. Sustainability. 13, 8, (2021).

https://doi.org/10.3390/su13084434.

32. Wang, Y. et al.: Acquiring logistics process intelligence: Methodology and an application for a Chinese bulk port. Expert Syst. Appl. 41, 1, (2014). https://doi.org/10.1016/j.eswa.2013.07.021.

33. Weijters, A.J.M.M., Van der Aalst, W.M.P.: Rediscovering workflow models from event-based data using little thumb. Integr. Comput. Aided. Eng. 10, 2, (2003). https://doi.org/10.3233/ica-2003-10205.

34. Wirth, R.: CRISP-DM : Towards a Standard Process Model for Data Mining. Proc. Fourth Int. Conf. Pract.

Appl. Knowl. Discov. Data Min. 24959, (2000).

35. Wynn, M.T. et al.: ProcessProfiler3D: A visualisation framework for log-based process performance comparison. Decis. Support Syst. 100, (2017). https://doi.org/10.1016/j.dss.2017.04.004.

APPENDIX

A. Technology Acceptance Model

Table 7. Scoring scale TAM

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Table 8. Technology Acceptance Model

Construct	1		2		3		4		5		6		7		8		Mean	
General Information																		
1. I have much experience with business process modelling in general.	4;3		4;3		4;5		2;1		5;3		4;4		2;2		3;2		3,19	
2. I have much experience with process mining in general.	2;1		3;3		3;4		2;1		4;4		4;4		2;2		2;1		2,63	
Usefulness																		
3. The information presented in this model is useful for my daily job.	4;1	5;2	1;2	3;4	3;5	4;4	1;2	2;3	2;1	3;2	2;2	4;3	1;1	2;2	1;2	3;4	1,94	3,13
	5;2	2;1	4;3	3;3	4;2	5;1	4;4	3;4	4;3	5;4	3;4	2;4	4;3	5;5	4;4	3;1	3,56	3,19
4. The model is suitable for gaining new insights about the business process.	3;2	4;3	1;1	2;3	2;3	3;4	1;1	5;3	1;1	3;3	1;1	5;2	2;2	3;3	2;1	4;3	1,56	3,31
	4;2	2;3	3;3	3;4	4;4	5;5	4;4	2;4	4;4	5;2	3;4	1;3	4;5	4;4	4;5	5;4	3,81	3,50
5. The model contains detailed information about the business process.	2;1	3;3	1;2	3;2	1;2	2;3	1;2	3;3	1;2	3;2	2;1	3;3	1;1	2;3	1;1	2;4	1,38	2,75
	4;4	5;4	4;4	5;5	4;4	5;5	4;5	4;3	4;4	5;5	4;4	4;5	4;4	5;5	4;5	5;5	4,13	4,69
6. The model helps forming an understanding of the business process in general.	3;2	4;3	2;1	4;4	3;5	3;4	1;1	4;3	5;1	4;4	2;2	5;4	2;3	3;4	2;2	4;5	2,31	3,88
	3;4	1;2	3;4	2;2	4;3	4;2	3;5	1;2	4;4	4;3	3;3	2;3	3;4	3;2	3;4	2;1	3,56	2,25
Ease of Use																		
7. The model is understandable when taking a first look at it.	4;2	4;2	3;4	3;3	2;4	3;2	5;4	4;3	5;4	5;4	4;5	4;5	5;5	4;4	4;4	3;4	4,00	3,56
	3;3	1;1	2;3	1;2	4;2	3;1	3;3	2;1	4;3	3;2	3;3	1;2	2;3	1;1	2;3	1;2	2,88	1,56
8. It is easy to learn understanding this model.	4;2	3;2	4;4	3;4	2;4	2;2	4;5	4;3	5;4	3;5	4;5	4;4	5;4	5;4	5;4	5;3	4,06	3,50
	3;3	1;1	3;3	2;2	4;2	3;1	3;3	2;2	2;4	2;3	3;4	1;2	3;3	2;2	3;3	1;1	3,06	1,75
9. It is easy to explain this model to other persons inside the organization.	4;2	4;4	2;3	4;3	2;3	3;2	5;5	4;3	5;3	3;4	4;5	4;4	5;5	4;4	5;2	4;5	3,75	3,69
	4;3	2;2	4;3	2;1	3;1	1;1	3;4	2;1	2;4	1;3	2;4	1;2	2;3	1;1	3;4	1;2	3,06	1,50
10. Someone without experience in process mining is able to understand this model.	5;2	3;3	2;4	2;4	2;2	3;2	5;5	4;4	2;4	2;4	4;3	4;4	4;5	4;4	4;3	4;4	3,50	3,44
	3;2	2;2	2;3	1;1	3;1	2;1	3;3	2;2	1;3	1;2	2;4	1;2	2;4	1;2	2;3	1;1	2,56	1,50
Intention																		
11. I will use the information obtained from this model in my daily job.	3;1	4;2	1;2	4;5	2;4	3;3	1;2	2;2	2;1	4;3	2;3	4;3	2;2	3;3	1;1	4;4	1,88	3,31
	4;2	2;1	3;4	2;2	4;2	4;1	4;5	3;3	5;3	4;4	3;4	1;4	4;4	4;3	4;3	2;2	3,63	2,63
12. This model helps me achieve my purpose inside the organization.	4;1	5;3	1;2	4;5	1;4	2;4	1;1	2;2	2;1	4;2	2;2	4;3	2;1	2;4	1;1	4;3	1,69	3,31
	5;2	3;1	4;5	3;3	3;3	4;3	4;5	3;3	5;3	4;4	3;4	1;3	4;4	4;3	4;3	3;2	3,81	2,94
Total	2,60	3,30	2,15	3,45	2,80	2,90	2,65	3,15	2,60	3,35	2,80	3,80	2,90	3,35	2,35	3,80	**2,61**	**3,39**
	3,25	1,95	3,35	2,45	3,05	2,85	3,80	2,45	3,50	3,30	3,35	2,25	3,45	2,90	3,50	2,25	**3,41**	**2,55**