# Utilization of time-driven activity-based costing and process simulation in cost management of organization

**Michal Halaska[1], Roman Sperka[2]**

[1] Silesian University in Opava, School of Business Administration in Karvina, Department of Business Economics and Management, Czech Republic, ORCID: 0000-0002-7086-9137, halaska@opf.slu.cz;

[2] Silesian University in Opava, School of Business Administration in Karvina, Department of Business Economics and Management, Czech Republic, ORCID: 0000-0003-2371-6450, sperka@opf.slu.cz.

***Abstract:*** *The deployment of information and communication technologies in organizations is on the rise. Many organizations consider the application of technologies to be a crucial key to improve their processes. However, traditional costing systems are not suitable for cost estimation of business processes due to the use of volume-based cost drivers, which are often not adequate for the structure of today's organizations. In this research, we present an overview of how the TDABC (time-driven activity-based costing) model can be combined with process mining and business process simulation for cost estimation of such processes. The objective of this paper is to use the cost dimension as a major attribute for the potential implementation of robotic process automation (RPA) in companies. However, information and communication technologies could be considered in general. We demonstrate our approach in a case study that takes advantage of a real-life event log containing transactional data representing the loan application process in an insurance company. The event log is analyzed and processed using process mining techniques. Based on the preprocessing, a simulation model representing the original loan application process is designed. The designed simulation model is then used for simulation of partial and full implementation of RPA through separate scenarios. Then, we add the cost dimension to the simulation by enriching the event log with cost data based on a formalized cost model. We show that even though partial implementation of RPA might not deliver significant increase in efficiency in the process, it might still represent significant cost savings.*

***Keywords:*** *Business processes, process mining, simulation, TDABC, RPA, automation.*

***JEL Classification:*** *M4, M1, C63.*

***APA Style Citation:*** Halaska, M., & Sperka, R. (2024). Utilization of time-driven activity-based costing and process simulation in cost management of organization. *E&M Economics and Management*, *27*(2), 50–68. https://doi.org/10.15240/tul/001/2024-5-007

***Early Access Publication Date:*** *April 10, 2024.*

## Introduction

For the organization to succeed in today's highly competitive markets, it is necessary to improve its methods and processes, lower costs, and continually increase quality. The main purpose of accounting is to facilitate economic resources. This is achieved by providing accurate information to operating management, owners, and creditors, for price setting, budgeting, planning, forecasting, and other kinds of decision-making (Sánchez & Batista, 2020; Zamrud & Abu, 2020). It can be used to determine which products are profitable, and identify waste and low value-adding activities at the operational level. This is especially important regarding the philosophy of present managers, which places more emphasis on customer satisfaction. In numerous businesses, expenses exhibit significant variations because of the diverse range of products or operations they engage in, necessitating the implementation of more advanced and adaptable costing systems.

Contemporary cost accounting systems should exhibit dynamism and flexibility, enabling the computation of various categories of cost objects, such as products, activities, distribution channels, and clients, among others. This should be done while considering the full spectrum of diversity and complexity inherent in modern production and business operations, such as high proportion of indirect costs shared between multiple cost objects, heterogeneity of cost drivers and related demand, cost-competitiveness, and availability of information systems gathering information about costs and cost drivers. Traditional-based costing was introduced when the overhead costs were significantly lower and direct labor accounted for a major portion of production cost (Pashkevich et al., 2023; Vedernikova et al., 2020). Traditional costing, based on averages, worked for simple, uniform products with mostly variable costs. But in today's dynamic manufacturing and services, overhead costs are substantial (Popesko, 2010). Traditional accounting assumed stable markets, long product cycles, and large production runs, which no longer apply due to automation, shorter product life, and greater variety. This shift has altered cost structures (Pashkevich et al., 2023; Popesko & Tučková, 2012). Traditional costing distorts product costs and cannot break them down by activities (Lu et al., 2017; Vedernikova et al., 2020), hindering decision-making. It offers little incentive to control overheads, overestimating high-volume and underestimating low-volume costs (Hadid & Hamdam, 2022).

Recent IT advancements significantly enhance cost data collection and communication in organizations, facilitating evidence-based cost analysis of business processes. Traditional costing systems, however, underutilize available data in modern organizations (Vedernikova et al., 2020; Zamrud & Abu, 2020). This study focuses on integrating TDABC (time-driven activity-based costing) with process mining techniques. It aims to incorporate cost considerations into implementing RPA (robotic process automation) and information technology. To achieve this objective, we postulated the following research questions (*RQ*).

*RQ1: How can the TDABC model integrated into the event log used for the utilization of process mining techniques?*

*RQ2: How can process-related cost information enrich the implementation of RPA?*

The paper is structured as follows: in the first section, we present TDABC systems. The second section introduces process mining. The third section presents the background and methodology of the research. The fourth section presents the case study and its results. The fifth section discusses the advantages and limitations of our approach. To conclude, we summarize our results and the perspectives for its further development.

## 1. Time-driven activity-based costing systems

The organizations that gain the greatest advantage from adopting activity-based costing (ABC) systems are those with high frequency of different cost objects, those with a large portion of indirect and supporting costs, and those with a great number of processes and activities (Popesko, 2010). The major difference between ABC and TDABC is that TDABC uses a time driver for each activity. TDABC involves estimating solely the unit cost of supplying capacity and the time required to perform a transaction or an activity (Cidav et al., 2020). TDABC allows usage of multiple time drivers to estimate the time spent on each activity which makes implementation, maintenance and updating of TDABC easier (Ghani et al., 2020). ABC systems often lead to oversimplification of activities as it might have to treat varying transaction times by treating each variant of the process as a distinct activity in the case of multi-driver activities, which creates difficulties in estimating the practical capacity for each sub-activity (Gervais et al., 2010). With TDABC, all we need to know is the quantity of resources required to produce one unit or one batch of the cost object and how much of the cost pool should be allocated to the cost object (Sánchez & Batista, 2020). TDABC similarly to ABC supports operational improvements as it provides more cost transparency than traditional costing systems; thus, providing relevant information for managerial decision-making (Pashkevich et al., 2023; Sachini et al., 2022). The model eliminates activity pools, thereby simplifying the allocation of resource costs without the need for activity pools (Zamrud & Abu, 2020). TDABC was reported to be more efficient and simpler than ABC (Keel et al., 2017). The use of multiple time drivers eliminates the need for subjective, costly, and time-consuming workshops, observations, interviews, and surveys, which might provide biased process views (Cidav et al., 2020; Rahman et al., 2019). Furthermore, there are usually fewer resource groups than activities, which reduces the measurement error (Gervais et al., 2010). At the operational level, there are proven advantages through waste reduction, reduction of redundant human resources, non-value adding steps and waiting times, resulting in a detailed understanding of the cost of processes (Keel et al., 2017; Zamrud & Abu, 2020).

Together with the possibility to use data from ERP and other systems, it makes it easier to model sub-activities and maintain the modelled system in the long run (Pashkevich et al., 2023; Sachini et al., 2022). The use of different information and automatic data capture systems is no longer a privilege of large companies. Information systems help collect data about companies' processes with minimum manual effort. In this way, real-process information can be linked to cost information, resulting in automatic cost accounting. Businesses have the option to utilize alternative data collection approaches, such as radio frequency identification or automated measurements, as well as direct observations. Although TDABC is much easier to implement than ABC, data collection and analysis are laborious tasks, which may disregard the use of TDABC. The steps regarding the implementation of TDABC are very similar throughout the literature (e.g., Cidav et al., 2020; Vedernikova et al., 2020). There may be differences based on specific domains; however, in this research, we are looking for a generic approach towards the implementation of TDABC. Thus, in this research, we use the following implementation approach of TDABC:

1. **Identify activities.** In this step, we pinpoint the activities for which we seek cost information. A crucial prerequisite for implementing TDABC is having a comprehensive understanding of all processes and activities involved (Ribadeneira et al., 2019). This step implicitly includes the identification of various resource groups used by each activity.
2. **Determine the capacity cost rate.** The calculation of the capacity cost rate for these activities is essential. The capacity cost rate (CCR) can be defined as follows:

$$CCR = \frac{Cost\ of\ capacity\ supplied}{Practical\ capacity\ of\ resources\ supplied} \quad (1)$$

The cost of supplied capacity pertains to the expenses associated with the resources utilized for carrying out activities, while practical capacity indicates the amount of time allocated to perform the activity.

3. **Estimate activity time.** It is necessary to estimate the amount of time required for the completion of a single unit of the activity.
4. **Determine the cost driver rate.** The cost driver rate for an activity in TDABC is straightforwardly determined by multiplying the CCR by the time estimate required for a single occurrence of that activity.
5. **Assign activity costs to cost objects.** In this step, the expenses are allocated to the target by calculating the cost by multiplying the actual volume of cost driver units by driver rate.

For a comprehensive understanding of data, both process and organizational data must be seamlessly integrated and analyzed. To create these integrated datasets, it is essential to link the organizational data connected to each activity, whether it is performed by humans or automated systems, back to the corresponding activity within a business process case.

## 2. Process mining

Process mining encompasses a range of techniques aimed at helping organizations comprehend, assess, and enhance their processes (Garcia et al., 2019). Consequently, process mining holds promise as a valuable tool for implementing TDABC systems and supporting cost management decision-making. The cornerstone of process mining lies in the analysis of event logs, which are extracted from diverse information systems utilized within organizations, offering invaluable insights into the execution of work processes. These logs must be sourced from operational systems such as customer relationship management, enterprise resource planning systems, or even embedded systems. Regardless of the type of process mining analysis, the log should contain the following essential information: i) each event contained in the log must be unique and must be sorted; ii) it is crucial to be able to differentiate between process instances; and iii) there must exist a function capable of assigning an activity name to each event.

Various types of PM analysis necessitate distinct supporting attributes. For example, constructing a social network from event logs requires the inclusion of resource information within the log. Process mining techniques usually do not consider the cost dimension of the analyzed processes. Therefore, typical event logs do not contain cost-related data. The conventional format for event logs used to be MXML. However, due to various limitations, a new standard event log format called XES was introduced. Nonetheless, there are alternative formats available, such as CSV files or software-specific FXL files, among others. Within this research, we utilize two fundamental areas of process mining: process discovery and conformance checking. The primary objective of process discovery is to identify pattern within the logs, from which a process model of the observed process is constructed. Presently, one of the most successful techniques for this purpose is called split miner (Augusto et al., 2018). It is important to note that none of these techniques guarantees that the resulting model precisely corresponds to the original process or fully represents the behavior observed in the data. Therefore, it is crucial to verify the quality of the discovered process model through conformance checking (Garcia et al., 2019).

## 3. Background and methodology

This section provides the formal preliminaries for our research. Through the formalization of event logs, we establish precise requirements based on those outlined in Section 2, without delving into specific syntax considerations. This formal representation serves as a foundation for querying the event log and as a starting point for subsequent analysis and reasoning. Moreover, it presents data and methodology of the case study used. Firstly, we formalize the event log and then we extend the log by parameters of the cost model.

**Definition 1 (event, attribute).** Let $E$ be the set of all possible event identifiers and $AN = \{a_1, a_2, \ldots, a_n\}$ be the set of all possible attribute names. For each attribute $a_i \in AN (1 \leq i \leq n)$, $D_{a_i}$ is the set of all possible values for the attribute $a_i$. *For* any event $e \in E$ and an attribute name $a \in AN$, we denote $\#_a(e) \in D_a$ as the attribute's value name $a$ for event $e$. For any event $e \in E$, we define the following standard attributes: $\#_{case}(e) \in D_{case}$ is the case identifier of $e$; $\#_{id}(e) \in D_{id}$ is the event identifier of $e$; $\#_{act}(e) \in D_{act}$ is the activity name of $e$; $\#_{res}(e) \in D_{res}$ is the resource that triggered the occurrence of $e$; $\#_{stime}(e) \in D_{stime}$ is the start timestamp of $e$; and $\#_{ctime}(e) \in D_{ctime}$ is the complete timestamp of $e$.

**Definition 2 (case, trace, event log).** Let $C$ be the set of all possible case identifiers. For any $c \in C$ and an attribute name $a \in AN$, we denote $\#_a(c) \in D_a$ as an attribute's value named $a$ for case $c$. We denote $E^*$ as the set of all finite sequences of events over $E$ where a finite sequence of length $n$ over $E$ is a mapping $\sigma \in \{1, \ldots, n\} \rightarrow E$ and is represented as $\sigma = \langle e_1, e_2, \ldots, e_n \rangle$ where $e_i = \sigma(i)$ for $1 \leq i \leq n$. We define the special attribute $\#_{trace}(c) \in E^*$ as representing the trace of case $c$, which consists of all events associated with $c$. An event log $L \subseteq E$ is a set of events. We denote $\hat{c} = \#_{trace}(c)$ as shorthand for referring to the trace of a case and further note that the ordering in a trace should respect timestamps, i.e., for any $c \in L$, $i$, $j$ such that $1 \leq i \leq j \leq |\hat{c}|$: $\#_{time}(\hat{c}(i)) \leq \#_{time}(\hat{c}(j))$.

TDABC focuses on organization's resources. For each resource cost pool, resource costs are allocated to cost objects using two sets of estimates: i) the capacity cost rate (Equation (1)); and ii) an estimate of time units required to perform a process, an activity, or a service. Based on the TDABC cost model, we have to extend the event log with additional attributes. Thus, we define the following attributes:

- $\#_{type}(e(j)) \in D_{type}$ is the activity type of event $j$ where $D_{type}$ is a finite set;
- $\#_{RP_i}(e(j)) \in D_{RP}$ is the $i^{th}$ resource pool of event $j$ where $D_{RP}$ is a finite set;
- $\#_{PC_{RP_i}}(e(j)) \in D_{PC} \subseteq R_{\geq 0}$ is the practical capacity of $i^{th}$ resource pool of event $j$;
- $\#_{CCS_{RP_i}}(e(j)) \in D_{CCS} \subseteq R_{\geq 0}$ is the cost of capacity supplied of $i^{th}$ resource pool of event $j$;

- $\#_{CCR_{RP_i}}(e(j)) \in D_{CCR} \subseteq R_{\geq 0}$ is the $i^{th}$ capacity cost rate of event *j*;
- $\#_{dur}(e(j)) \in D_{dur} \subseteq R_{\geq 0}$ is the duration of event *j;*
- $\#_{odur}(e(i), e(j)) \in D_{odur} \subseteq R_{\geq 0}$ is the overlapping duration of events *i* and *j*.

### 3.1 TDABC costing model

Practical capacity, time and cost represent crucial components within the TDABC system, wielding substantial influence on its outcomes. The inaccurate estimation of these elements can result in significant deviation from the actual values when calculating the costs associated with products or services. The TDABC costing model can be described as follows: Let *L* be an event log and parameters of the cost model be represented as attributes of the event log as defined above.

In step 1, all activities related to the modelled process and related cost objects, and resources used by these activities and their costs have to be identified. These resources are then grouped into resource pools based on resource cost drivers. Based on defined resource pools, the cost of capacity supplied for each resource pool is estimated.

In step 2, the capacity cost rate of each resource pool is estimated based on the cost of capacity supplied and practical capacity. Practical capacity is estimated as the quantity of resources (typically, personnel or equipment) that perform activities. It refers to the actual capacity of resources, as opposed to the amount of time theoretically available for performing the activity. Thus, the practical capacity for each activity type is estimated as follows:

$$\#_{PC_{RP_i}}(e(j)) = \sum_{j=1}^{k} \#_{dur}(e(j)), \quad \text{where } \#_{type}(e(j)) = b_i \tag{2}$$

Each activity type has assigned exactly one resource pool. Activities are classified based on activity type to differentiate between potentially different outputs of activities recorded in the log, which might consume different resources. Another option is to split such activity into several activities; however, this would add to the complexity of the process model, which is not desirable. Furthermore, *k* represents the number of events in the log *L*. Practical capacity is measured in units of time. Using Equation (2), the capacity cost rate is defined as follows:

$$\#_{CCR_{RP_i}}(e(j)) = \frac{\#_{CCS_{RP_i}}(e(j))}{\#_{PC_{RP_i}}(e(j))}, \quad \text{where } \#_{type}(e(j)) = b_i \tag{3}$$

In step 3, management estimates the amount of time it takes to a complete single unit of the activity. Usually, activity times are estimated involving different techniques like, e.g., direct observations, employee surveys. Based on activity drivers and time estimates, time equations are formulated which then determine durations of instances of activities. In our approach, we use timestamps recorded by the information systems in the event logs. Activity duration is estimated using the start timestamp and end timestamp of the event. Thus, we define the duration of the event as follows:

$$\#_{dur}(e(j)) = \#_{ctime}(e(j)) - \#_{stime}(e(j)) \tag{4}$$

In step 4, the cost driver rate for an activity in TDABC is defined as the product of the capacity cost rate and the time estimate for a single unit of the activity. However, since we are using exact times at the event level, the cost driver rate is not useful. Thus, we define the activity cost (*AC*) as follows:

$$AC = \sum_{j=1}^{k} \#_{dur}(e(j)) * \#_{CCR_{RP_i}}, \quad \text{where } \#_{type}(e(j)) = b_i \tag{5}$$

In step 5, we assign activity costs to the cost object. The costs are assigned to the cost object by summing over entire event log. Thus, we define the total cost of the process as follows:

$$TC = \sum_{j=1}^{k} \sum_{i=1}^{m} \#_{dur}(e(j)) * \#_{CCR_{RP_i}} \tag{6}$$

### 3.2 Data

The research utilizes a publicly available event log (van Dongen, 2012) to study the execution of the loan application process in the year 2011. This event log comprises 13,087 process instances, or cases, involving a total of 262,000 events. Each event within the log is associated with one of the 36 activity names. The log covers a reference period from October 1, 2011, to March 14, 2012, with an average case duration of 8.6 days. Events in the log are categorized into three types, each denoted by an event name beginning with either A, O, or W. Specifically, A events relate to applications, O events pertain to offers sent to customers, and W events concern the processing of work items associated with applications. Each event recorded in the log comprises nine attributes, including “Case ID” for case identification, “Activity” for identifying activities within cases, “Resources” for identifying the resources responsible for event execution, and “Start timestamp” and “Complete timestamp” for pinpointing even occurrences. The overall workflow of the process unfolds as follows. Following the application submission, a subset undergoes scrutiny for fraudulent behavior, while the rest are examined for completeness. Subsequently, applications are pre-accepted and processes. Some applications are canceled, while offers are dispatched to the remaining customers, followed by customer contact. If the customer accepts the offer, the application undergoes assessment, leading to loan approval. In certain cases, further customer interaction might be necessary post-assessment to finalize the application. It is important to note that while the event log contains real-life transactional data, it lacks the attributes used by the presented costing model. Therefore, for the purpose of this case study, we augmented the event logs by incorporating artificial cost attributes.

## 3.3 Methodology

### Data preparation

At this stage, it becomes imperative to prepare the log for application of individual process mining techniques. This involves extracting logs from various database sources. Process mining techniques are typically executed within software tools that operate with specific data file formats such as CSV, XES, XML, and MXML. Fortunately, the log pertaining to the loan application process was already available in CSV format. An essential step in this process involved verifying whether all events in the log contained the fundamental attributes required in the appropriate format. Events and associated cases that either lacked the necessary attributes or did not adhere to the required formats were either modified to comply with these criteria or excluded from the dataset. Any missing values were addressed through a similar approach. The outcome of this preparation stage is the refined event log, which is subsequently employed in the process discovery phase to uncover the process model of the loan application process.

### Process discovery and analysis

When implementing a TDABC costing system, the identification of activities within the process and gaining insights from various process perspectives is crucial. In this context, process mining, as a data-centric approach, offers a significant advantage over other commonly used business process management practices like workshops or interviews. These traditional methods may not provide a comprehensive understanding of the entire process, especially when process owners are not intimately familiar with every aspect. To achieve this, we employed process discovery techniques within the Apromore process mining tool, allowing us to generate an executable process model. Apromore’s process discovery relies on the split miner technique (Augusto et al., 2018), which excels in terms of fitness, precision and *F*-score when compared to other process discovery methods. Noise within the log is effectively filtered out using integrated nodes, arcs, and parallelism filters. The node filter screens activities based on their frequency of occurrence, the arcs filter operates based on the frequency of arc occurrences, and the parallelism filter enables adjustment for parallelism, such as AND and OR gates, discovered in the process. The quality of the process model derived through Apromore is assessed using fitness, precision, and *F*-score metrics (Buijs et al., 2012). The output of this process discovery stage yields the process model for the loan application process, forming the foundation upon which the subsequent simulation is constructed.

### Process simulation model

During this stage, we design the simulation model, which serves as a tool for assessing the impact of operation changes on the cost perspective of the process. For our simulation, we adopted the approach presented by Măruşter and van Beest (2009) as well as Rozinat et al. (2009). Initially, we adjusted the discovered process model for simulation purposes using filter nodes, arcs, and parallelism. The aim was to simplify the model while retaining a maximum number of activities from the log and ensuring an acceptable level of model quality. Decision points within the model, such as XOR gates, were simulated as percentages based on the frequency of outgoing arcs. In essence, these probabilities were determined as a mathematical ratio of cases affected by the selected path. The business instance for the simulation was directly derived from the original log, which contained 13,087 cases. However, the simulation tool we used, BIMP, had a maximum limit of 10,000 process instances. Therefore, we worked with 10,000 cases as it provided a sufficient representation of the process.

The time-related parameters in the simulation model were also based on the original logs and consisted of three components: the arrival of new cases, processing times, and waiting times. The arrival of new cases was modeled as a Poisson process, with arrival times following a Poisson distribution estimated from the original log. For processing times, we employed various probability distributions. The selection of the appropriate distribution was based on its fit to the data using Kolmogorov-Smirnov statistics, Cramer-von Mises statistics, and standard error. When distribution has similar Kolmogorov-Smirnov statistics, the one with lower Cramer-von Mises statistics was chosen. In cases where the best-fitting distribution had a high standard error leading to a significant deviation in the mean value, the standard error was considered during the selection process. The same probability distributions and selection procedures were applied to estimate waiting times. In the original log, activities were categorized into two states: "Start" and "Complete" with processing times defined as difference between the complete timestamp and start timestamp for each activity. Similarly, waiting times represented the time interval between the start timestamp of an activity and the complete timestamp of the preceding activity. We treated activities with instant processing times in the original log, where start and complete timestamps were the same as having instant processing times in the simulation model. Activities with instant waiting times in the original log, signifying very short waiting times on the order of milliseconds, were approximated as having a waiting time value of 0.001 milliseconds.

The simulation model's organizational perspective stems from the original log data. This perspective primarily focuses on resource allocation for various activities. According to the data, most resources are involved in a wide range of activities. The variation lies in the extent to which each resource engages in each specific activity, indicating how frequently the resource carries out that activity throughout the log. Activities that are instantaneously processed are attributed to the resource labeled as "System." To allocate distinct resource types to each non-instantaneous activity, we employ the *k*-means clustering technique. This clustering approach hinges on the distinctive profile of each resource. The resource profile is represented as a one-dimensional vector, with each element indicating how many times the resource executed each non-instantaneous activity. Determining the optimal number of clusters for *k*-means clustering is accomplished through the utilization of several methods, including the elbow method, the silhoutte method, and gap statistics. Each activity is then assigned to a specific cluster based on the cluster with the highest cumulative number of executions of that activity by its constituent resources (Equation (7)).

$$\max_{\mathrm{n}} \sum_i \mathrm{e}_i^n \tag{7}$$

where: $n$ – the number of clusters; $i$ – the number of resources within the cluster; $\mathrm{e}_i^n$ – the number of times activity a was executed by resources $i$ belonging to cluster $n$.

The reliability of results obtained through simulation increases as the approximation becomes closer to reality. To achieve a simulation that closely mirrors real-world data, the following similarity indicators are employed:

- Process flow and semantics – both models should exhibit identical process flow, representing the sequence of activities and BPMN constructs. Furthermore, in terms of semantics, it is crucial that the labels assigned to the activity are the same;
- Throughput times – the throughput times for each activity in the discovered process model should closely align with those in the simulated process model for comparison;
- Bottlenecks – the location and severity of bottlenecks in the process model should closely match those in the simulation model for comparison.

### What-if analysis

At this stage, we present the evaluation of the cost estimation of the business process. It is based on a combination of TDABC, process mining, and business process simulation. The evaluated scenario focuses on the deployment of RPA in the loan application process. RPA technology is employed to automate business processes that were previously carried out by employees. As a result, this technology allows employees to engage in more complex tasks, thereby delivering greater value to the organization. RPA is best utilized for automation of rule-based, highly frequent, and repetitive tasks that are prone to error (Syed et al., 2020; Šperka & Halaška, 2023). Based on the features of RPA, we identified several activities within the loan application process suitable for the deployment of RPA. We implement these changes into simulation model and demonstrate our approach towards cost estimation using TDABC model.

# 4. Case study

## 4.1 Data preparation

In the initial stage, it is imperative to prepare the logs to facilitate the application of specific process mining techniques. For our study involving the loan application process, we had access to logs available in both XES and CSV formats. We opted for the CSV format due to its user-friendliness and accessibility, especially for individuals without prior expertise. Unlike XES, which relies on XML with a predefined syntax, CSV offers a more intuitive and straightforward structure. We conducted a thorough review to ensure that all events in the log adhered to the essential prerequisites. This included confirming the presence of unique case IDs, ensuring that each event was associated with an activity name, verifying that timestamps followed the “dd.mm.yyyy hh:mm:ss” format, and confirming that events were appropriately linked to designated resources. Any events or related cases lacking these required attributes or not conforming to the specified formats were either modified to meet these criteria or excluded from the dataset. We adopted a similar approach in handling missing values. It is worth noting that our primary focus was exclusively on the specified attributes, namely case IDs, activities, timestamps, and resources. Other attributes were disregarded. Consequently, no cases or events were omitted from the log based on these auxiliary attributes. The outcome of data preparation culminated in the creation of an event log, which served as the foundation for subsequent stages involving process discovery and analysis within the context of the loan application process.

**Process discovery and analysis**

The 2011 log comprises a total of 13,087 instances, encompassing a substantial 262,000

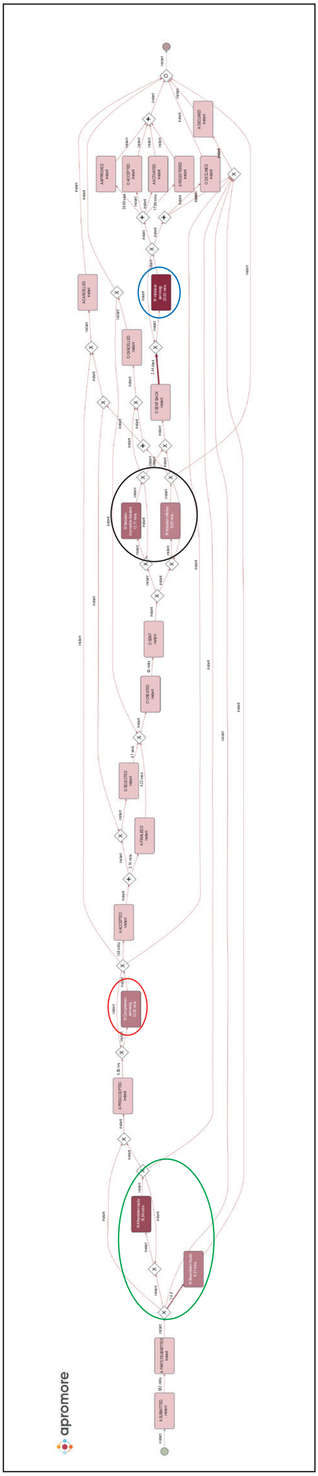

**Fig. 1:** BPMN process model representing the loan application process from 2011 using Apromore (values of filters nodes, arcs and parallelism set up to 100, 60, 100)

Note: Fig. 1 is for illustrative purposes only to showcase the process flow of the discovered process.

Source: own

individual events. Within this log, 23 distinct activity names have been attributed to all recorded events. It is essential to highlight that the loan application process exhibits notably complex and unstructured overall behavior. Thus, the discovered process has to be simplified to obtain a process model suitable for simulation purposes (Fig. 1). Of the 23 activities, seven activities are being done manually, and the rest of them are automated. This means that the manual activities have processing time, while the rest of the activities are being done almost instantly. Thus, in the cost analysis of the process, we will focus on the manual activities because the automated activities do not provide much information at the case level. Furthermore, we excluded activity "W_Change contract details" from the cost analysis to make the simulation model simpler as it appears in the entire log only 12 times. Similarly, we excluded loops of length one from the cost analysis because they greatly increase the number of process variants.

In Fig. 1, we can observe specific activities highlighted within colored ovals:

- Green oval – within this green oval, we find the activities "W_Assess potential fraud" and "W_Handle leads." The "W_Assess potential fraud" activity involves the valuation of loan applications for potential fraud indicators. Meanwhile, "W_Handle leads" is associated with the processing of incomplete initial application submissions;
- Red oval – in the red oval, we highlight the activity labeled "W_Call incomplete files." This activity pertains to the process of finalizing pre-accepted applications, particularly those that were initially incomplete;
- Black oval – inside the black oval, we find activities "W_Call after offers" and "W_Complete application." "W_Call after offers" corresponds to the stage where an offer is extended to a qualified applicant. Concurrently, "W_Complete application" involves gathering additional information during the assessment phase of the application.

Furthermore, it is worth noting that the "W_Validate application" activity serves as an essential component within the application review process.

### Process simulation

The first step involves verifying the adherence to the indicators used to assess the suitability of the simulation model, as discussed in Section 3.3 – Process simulation. It is important to note that the process flow and semantics of both the discovered and simulation models closely align, except for the inclusion of initial and final activities in the simulation model, which are artificially introduced but do not impact the core process logic, performance, or bottlenecks. They are represented as instant activities with immediate processing and waiting times, as depicted in Fig. 2 (highlighted by orange ovals). The design of the simulation model was based on the discovered BPMN process model. The discovered process model consists of 23 activities, while the simulation model encompasses 26 activities. Fig. 2 visually illustrates the process flows and bottlenecks in both the discovered and simulation models. Key activities are highlighted within colored ovals: "W_Handle leads" and "W_Assess potential fraud" in red, "W_Call incomplete files" in green, "W_Call after offers" in blue, and "W_Applications assessment" and "W_Validate application" in black. The BPMN model, serving as the foundation for the simulation model, exhibits the following quality values: a fitness score of 0.75, precision of 0.76, and *F*-score of 0.76. It is worth noting that the model's lower quality rating is primarily due to the absence of length 1 loops, which are not considered in this model. Length 1 loops can also adversely affect other discovery techniques. Nevertheless, the quality of the model remains satisfactory for its intended purpose.

Tab. 1 provides a comparative analysis of the bottlenecks between the discovered model and the simulation model. These values represent the averages of process and waiting times obtained from fifteen simulation runs. It is noteworthy that the severity of bottlenecks in the simulation model closely mirrors the situation observed in the discovered model. Based on the information presented in Fig. 2 and Tab. 1, it can be confidently asserted that the assessment indicators for the simulation model are satisfactory. Specifically, the alignment in terms of process flow, semantics, process throughput, and bottlenecks between both models is deemed adequate.

### What-if analysis

The simulation model is built on real-life transactional data contained in the event log. Data required for the cost perspective are

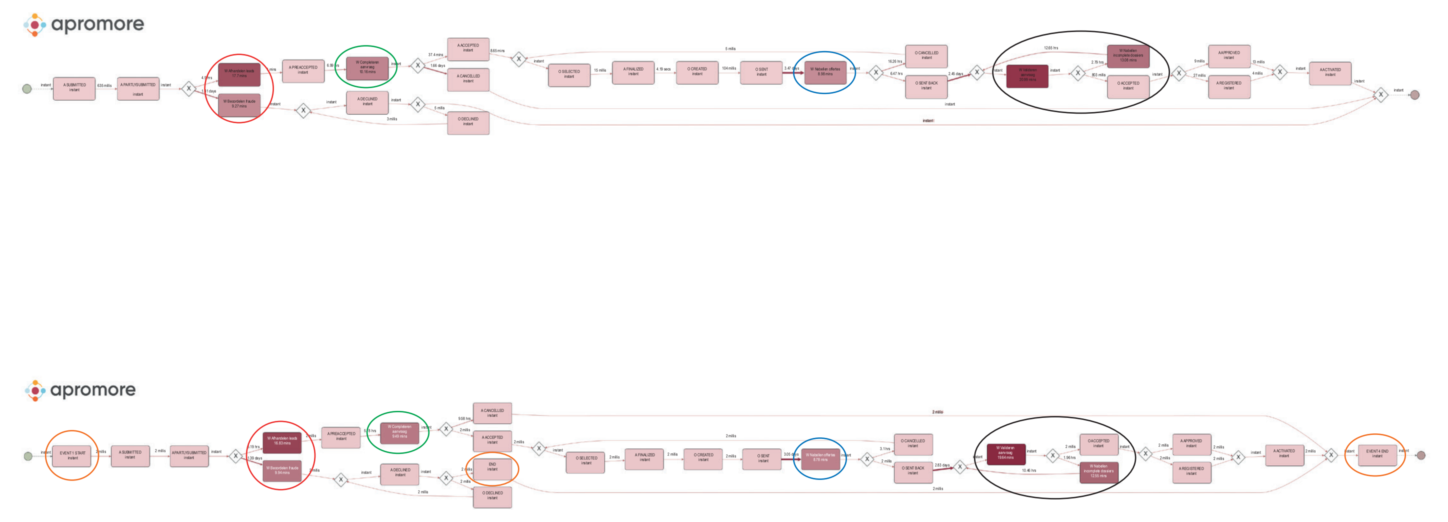

**Fig. 2:** Process flow of discovered (upper part) and simulation (bottom part) models

Note: Fig. 2 is for illustrative purposes only to showcase the process flow of the discovered process and simulation model.

Source: own

**Tab. 1:** **Processing and waiting times of activities in the original and simulation log**

| | | W_Handle leads | W_Assess potential fraud | W_Call incomplete files |
|---|---|---|---|---|
| **Frequency of occurrence in the original log** | | 4,755 | 108 | 7,367 |
| **Original log** | Processing times | 16.94 min | 9.27 min | 10.02 min |
| | Waiting times | 5.15 h | 1.40 d | 5.85 h |
| **Average of simulation logs** | Processing times | 16.94 min | 9.87 min | 9.47 min |
| | Waiting times | 5.20 h | 1.38 d | 5.74 h |
| **Std. dev. of simulation logs** | Processing times | 0.36 min | 0.79 min | 0.12 min |
| | Waiting times | 0.06 h | 0.13 d | 0.08 h |
| | | **W_Call after offers** | **W_Complete application** | **W_Validate application** |
| **Frequency of occurrence in the original log** | | 5,011 | 1,647 | 3,210 |
| **Original log** | Processing times | 9.08 min | 12.71 min | 20.92 min |
| | Waiting times | 2.83 d | 1.78 h | 2.44 d |
| **Average of simulation logs** | Processing times | 8.76 min | 12.83 min | 19.63 min |
| | Waiting times | 3.06 d | 1.70 h | 2.82 d |
| **Std. dev. of simulation logs** | Processing times | 0.10 min | 0.16 min | 0.20 min |
| | Waiting times | 0.03 d | 0.03 h | 0.01 d |

Source: own

added arbitrarily to demonstrate the possible implementation of our approach. For simplicity, Tab. 2 presents the $i^{th}$ capacity cost rate (CCR) of activity type *a*. The cost model is based on Section 3 and is the same for both the discovered and simulation models due to the focus on implementation of RPA technology within the loan application process. In our cost model, we assume that each activity has exactly one resource pool. However, based on attributes defined in Section 3, it is possible to assign more than one resource pool to each activity type, which would determine different capacity cost rates for each activity type. Furthermore, the cost model does not take into consideration direct costs, as they can be directly allocated to cost objects. It was not clear how to handle automated activities occurring in the log. Automated activities are being processed almost instantaneously in the 2011 log. Thus, it is not useful to use time drivers for such activities but rather a volume-based driver. However, it is again straightforward to extend the cost model presented to include volume-based drivers. The capacity cost rate is expressed in monetary units per minute, the average activity duration is in minutes, and the average activity cost per event is in monetary units per event. The CCR is arbitrary to demonstrate the case study and implications of the integration of TDABC and process mining techniques.

Tab. 3 provides an overview of the impact of RPA deployment on the performance of the discovered process. Achieving full automation entails the elimination of processing and waiting times associated with automated activities. Tab. 3 outlines six distinct scenarios, each focused on selected activities. The results for each scenario are calculated as the averages of data obtained from fifteen simulation runs. In the “Activity” column of Tab. 3, one finds the specific activity targeted for automation in each scenario. The “Average case duration” indicates the time required to complete all activities within a case, measured in days. The “Reduction of average case duration” represents

**Tab. 2:** **Capacity cost rates and cost characteristics at activity level**

| Activity | *RP* | CCR | Average activity duration | Average activity cost per event |
|---|---|---|---|---|
| **W_Handle leads** | $RP_{HL}$ | 0.09 | 16.943 | 1.525 |
| **W_Assess potential fraud** | $RP_{APF}$ | 0.15 | 9.883 | 1.482 |
| **W_Call incomplete files** | $RP_{CIF}$ | 0.10 | 9.493 | 0.949 |
| **W_Call after offers** | $RP_{CAO}$ | 0.10 | 8.759 | 0.876 |
| **W_Complete application** | $RP_{CA}$ | 0.03 | 12.833 | 0.385 |
| **W_Validate application** | $RP_{VA}$ | 0.12 | 19.627 | 2.355 |

Source: own

the percentage reduction in the average case duration after RPA deployment in comparison to the simulation model. “Workload reduction” quantifies the time savings realized through automation via RPA, measured in hours, and is directly related to the processing times of the activities involved. Similarly, “Potential cost savings” signifies the potential cost savings achievable through RPA deployment, also measured in hours, and relates to the processing times of activities within the log. These metrics offer valuable insights into the efficiency gains and cost reductions associated with the integration of RPA into the process.

Tab. 3 illustrates the impact of introducing RPA into the simulation model across various. Statistically significant alterations, denoted by “*”, in average case duration at a significance level of α = 5% are highlighted. It is worth noting that full automation exerts a statistically significant impact on the average case duration in all scenarios except for the activity “W_Assess potential fraud.” The most substantial changes in average case duration were observed in activities such as “W_Call after offers,” “W_Validate application,” and “W_Handle leads,” where reductions of 51.4737%, 38.2296%, and 3.8625% were achieved, respectively. In terms of workload

**Tab. 3:** **Efficiency and cost dimensions of simulations of to-be process model after RPA deployment – full automation (FA) and partial automation (PA)**

| Activity | Average case duration (FA) | Average case duration (PA) | Reduction of average case duration (FA) | Reduction of average case duration (PA) | Reduction of workload (FA and PA) | Potential cost savings (FA and PA) |
|---|---|---|---|---|---|---|
| **W_Call incomplete files** | 6.6733* | 6.9067 | 3.5831 | 0.2119 | 1,562.23 | 15,092.98 |
| **W_Validate application** | 4.2753* | 6.9053 | 38.2296 | 0.2312 | 2,664.08 | 151.21 |
| **W_Call after offers** | 3.3587* | 6.9133 | 51.4737 | 0.1156 | 1,188.39 | 9,396.59 |
| **W_Assess potential fraud** | 6.9033 | 6.9000 | 0.2601 | 0.3082 | 16.78 | 10,157.72 |
| **W_Complete application** | 6.7949* | 6.8714* | 1.8267 | 0.7214 | 997.22 | 4,207.31 |
| **W_Handle leads** | 6.6540* | 6.8688* | 3.8625 | 0.7590 | 2,794.32 | 44,918.99 |

Note: *denotes statistically significant results at a significance level of α = 5%.

Source: own

reduction, the implementation of RPA yielded the most favorable outcomes for activities including “W_Handle leads,” “W_Validate application,” and “W_Call incomplete files.” These findings underline the notable efficiency improvements associated with the deployment of RPA in these specific areas. Partial automation is specified by eliminating the processing times of automated activity. Partial automation significantly impacts average case duration in only two scenarios: “W_Complete application” and “W_Handle leads.” Moreover, the reduction of the average duration of the case is less than 0.7590% in all scenarios in the case of partial automation. However, due to the reduction of workload in the case of full automation and partial automation, one can see significant potential cost savings in both cases, even though partial automation offers much less savings with regard to the time efficiency dimension.

Moreover, there are several data quality challenges that directly affect design of cost model while using event logs and exact times that need to be kept in mind:

- Overlapping activities – overlapping activities are activities that are performed by the same resource during overlapping time intervals;
- Missing activities, events, and resources – sometimes missing, incorrect, or insufficient information might be recorded in the event log;
- Missing timestamps – sometimes, timestamps might not be recorded at all for one or more events or activities. Some systems also utilize different event lifecycles, which significantly impact the possible use of the cost model within the event log (Halaška, 2021).

In our case study, missing attributes (e.g., activities, events, resources, and timestamps) were dealt with in the data preparation phase by omitting such events from the log. This is acceptable for our case study; however, it is not for a real-life setting, as it disturbs cost allocation. In the 2011 event log, there were no missing attributes among activities, start, and complete timestamps. There were 8,141 missing resources of 164,509 observations, which occurred as follows: “W_Handle leads” with 817 occurrences, “W_Call incomplete files” with 3,098 occurrences, “W_Complete application” with 972 occurrences, “W_Call after offers” with 3,210 occurrences and “W_Validate application” with 44 occurrences. Another problem with TDABC and the use of exact times through process mining is that people are usually involved

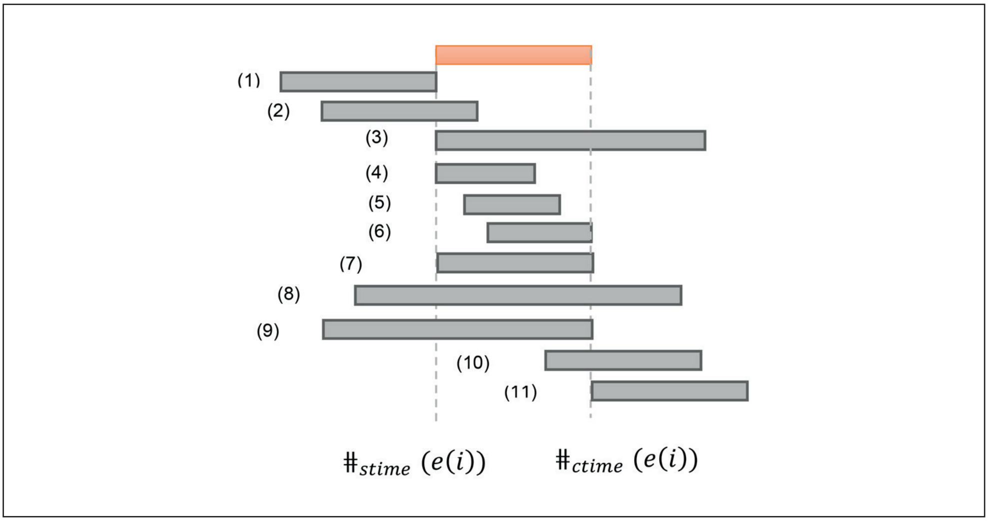

**Fig. 3:** **Overlapping time duration patterns of activities**

Source: own

in multiple processes. Thus, resources do not dedicate all their time to one task and may divide time simultaneously between different tasks. This is usually done based on priorities and workload. We show this problem and the impact of inaccurate cost allocation.

Tab. 4 presents inaccurate cost allocation due to the participation of multitasking based on specific resources. Since a total of 68 resources are used in the process, the 10 most utilized were selected to demonstrate the problem of cost allocation based on multitasking of resources. Multitasking was determined based on the overlapping time durations of activities performed by a specific resource (Fig. 3). For estimation of inaccurate cost allocation, we consider the following resources {10138,11169,10861,11181,10972,10609,11189,10913,11119,11180} $\in D_{resM} \subset D_{res}$, and activities $\{W\_Handle\ leads, W\_Call\ incomplete\ files,\ W\_Call\ after\ offers,\ W\_Complete\ application,\ W\_Assess\ potential\ fraud,\ W\_Validate\ application\} \in D_{actM} \subset D_{act}$. For each resource from $D_{resM}$, we firstly identified events with overlapping time durations based on the following condition: $\#_{ctime}\ (e(i)) \geq \#_{stime}\ (e(j)) \wedge \#_{stime}\ (e(i)) \leq \#_{ctime}\ (e(j))$. For each resource, only activities from the $D_{actM}$ were considered. Based on Fig. 3, we define the overlapping time duration for events *i* and *j* as follows:

For Patterns (2) and (9), we define: $\#_{odur}\ (e(i), e(j)) = \#_{ctime}\ (e(j)) - \#_{stime}\ (e(i))$;
For Patterns (3) and (10), we define: $\#_{odur}\ (e(i), e(j)) = \#_{ctime}\ (e(i)) - \#_{stime}\ (e(j))$;
For Patterns (4–7), we define $\#_{odur}\ (e(i), e(j)) = \#_{ctime}\ (e(j)) - \#_{stime}\ (e(j))$;
For Pattern (8), we define $\#_{odur}\ (e(i), e(j)) = \#_{ctime}\ (e(i)) - \#_{stime}\ (e(i))$.

Thus, overlapping time duration ($OTD_{resource \in D_{resM}}$) in minutes for a specific resource from Tab. 4 is derived based on Equation (8) and the overlapping time duration pattern in Fig. 3:

$$OTD_{resource \in D_{resM}} = \\ = \sum_{\forall e(i),\, e(j)\, \in\, L:\ \#_{res}(e(i)) = resource\ \wedge\ \#_{res}(e(j)) = resource\ \wedge\ \#_{act}(e(i)) \in D_{actM}\ \wedge\ \#_{act}(e(j)) \in D_{actM}} \#_{odur}\ (e(i), e(j)) \tag{8}$$

Overall duration in Tab. 5 is the duration that the resource spends working on activities from set $D_{actM}$ in minutes. Inaccurate allocation of costs ($IAC_{resource \in D_{resM}}$) in Tab. 5 is derived based on Equation (9):

$$OTD_{resource \in D_{resM}} = \\ = \sum_{\forall e(i),\, e(j)\, \in\, L:\ \#_{res}(e(i)) = resource\ \wedge\ \#_{res}(e(j)) = resource\ \wedge\ \#_{act}(e(i)) \in D_{actM}\ \wedge\ \#_{act}(e(j)) \in D_{actM}} (\#_{odur}(e(i), e(j)) * \\ * \#_{CCR_{resource}}(e(i)) + \#_{odur}(e(i), e(j)) * \#_{CCR_{resource}}(e(j))) \tag{9}$$

**Tab. 4:** **Inaccurate cost allocation due to participation in multitasking – Part 1**

| Resource | Overlapping time duration | Overall duration | Overlapping time duration to overall duration ratio | Inaccurate allocation of costs |
|---|---|---|---|---|
| **10138** | 64.24 | 26,490.21 | 0.2425 | 13.0807 |
| **11169** | 35986.82 | 75,856.05 | 47.4409 | 6,713.3370 |
| **10861** | 2119.24 | 16,175.06 | 13.5965 | 397.6516 |
| **11181** | 28.09 | 12,290.41 | 0.2285 | 4.6120 |
| **10972** | 24.05 | 41,232.85 | 0.0583 | 4.9451 |
| **10609** | 210.70 | 37,664.55 | 0.5594 | 47.0223 |
| **11189** | 496.21 | 14,255.58 | 3.2914 | 83.5944 |
| **10913** | 271.92 | 17,499.26 | 1.5538 | 40.2051 |

**Tab. 4:** **Inaccurate cost allocation due to participation in multitasking – Part 2**

| Resource | Overlapping time duration | Overall duration | Overlapping time duration to overall duration ratio | Inaccurate allocation of costs |
|---|---|---|---|---|
| **11119** | 303.99 | 11,742.24 | 2.5889 | 49.2679 |
| **11180** | 137.29 | 8,175.01 | 1.6789 | 23.4699 |

Source: own

**Tab. 5:** **The cost share of each activity on inaccurate cost allocation of resources**

| Activity | Resource | | | | |
|---|---|---|---|---|---|
| | **10138** | **11169** | **10861** | **11181** | **10972** |
| **W_Handle leads** | 0.00 | 1,084.51 | 328.15 | 0.00 | 0.00 |
| **W_Call incomplete files** | 0.00 | 5,022.45 | 16.35 | 2.17 | 0.00 |
| **W_Call after offers** | 0.02 | 424.08 | 44.36 | 2.01 | 0.00 |
| **W_Complete application** | 0.78 | 157.56 | 2.88 | 0.43 | 0.28 |
| **W_Assess potential fraud** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **W_Validate application** | 12.28 | 24.73 | 5.92 | 0.00 | 4.67 |
| | **10609** | **11189** | **10913** | **11119** | **11180** |
| **W_Handle leads** | 0.89 | 14.36 | 2.85 | 0.61 | 1.95 |
| **W_Call incomplete files** | 0.04 | 40.16 | 15.09 | 29.73 | 13.07 |
| **W_Call after offers** | 0.02 | 25.37 | 16.33 | 13.79 | 6.82 |
| **W_Complete application** | 1.08 | 3.71 | 5.94 | 4.93 | 1.61 |
| **W_Assess potential fraud** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **W_Validate application** | 44.10 | 0.00 | 0.00 | 0.21 | 0.01 |

Source: own

Since it was not clear to which of the overlapping durations the cost should be assigned, it is counted twice for each event with an appropriate capacity cost rate.

Tab. 5 shows the cost share of each activity in inaccurate cost allocation for each resource. If we compare it to Tab. 3, we can see that, for example, the potential time savings for the implementation of RPA in the activity “W_Call incomplete files” is 15,092.98 while inaccurate cost allocation within this activity is 5,022.45 by resource 11169. Similarly, potential time savings for activity “W_Handle leads” are 44,918.99, while inaccurate cost allocation within this activity is 1,084.51.

## 5. Discussion

There are three reasons for the extension of the event log (data model) instead of the process model. First, it is possible to use exact times instead of time equations. Setting up time equations requires deep knowledge of process variants and their corresponding drivers. However, the 2011 log contains more than 4,200 different process variants. This is not unusual when it comes to more complex processes with a higher number of activities. Furthermore, time equations are linear; however, activity durations might not always be well approximated using linear time equations. The duration of resource execution is not constant and should follow a probabilistic distribution. People also do not work at a constant speed. There are several studies suggesting a relationship between workload and performance (Nakatumba et al., 2012; Wickens et al., 2015). It is also not unusual to let work items related to the same task accumulate and then process all of them in one batch. The use of time equations also assumes, to a certain degree, stability of the process and organization and that neither of them changes over a certain period. However, if the flow of times becomes too long and work is accumulating. Resources may decide to skip certain activities or additional resources may be mobilized. To maintain simplicity, at a certain level, resources are treated as undifferentiated entities grouped into a resource pool with undifferentiated performance, where the processing times of an activity do not depend on the resources that perform it. The use of exact times found in the event log allows to treat resources individually, and the performance of each resource is independent of that of the other resource.

Second, extension of event log allows to stay at different levels of the analyzed process. TDABC considers non-financial measures for operational control and improvement. Thus, the TDABC covers both product and process levels for decision-making. That is, log, trace, activity, and event level. As was shown in the case study, it is possible to integrate TDABC with process mining techniques. Process mining techniques are applicable in the construction of the TDABC model, such as identifying activities and resources involved in the process. It can also assist in the automation of implementation. Maintenance and update together with the cost analysis process from both perspectives, product, and process perspective, as process mining would be involved in process mapping, measurement and cost allocation. Not all resources are identifiable using process mining; however, it is possible to extend event logs with data required for building a TDABC cost model. Such a system can also be used to assess the cost dimension of the implementation of different technologies, such as RPA, which was shown in the case study. The cost dimension in such cases has to be assessed from the process perspective. Process mining is dependent on the quality of data, and as such a costing system based on process mining would be as well. However, today companies collect large volumes of data related to their processes.

Third, process mining is well-suited for building simulation models. This is another advantage of the integration of TDBAC and process mining, since TDABC allows for the simulation of costs. This was also shown in the case study. The cost dimension is usually omitted from the analysis of business processes. The integration of process mining and TDABC has a huge potential for managing costs in a structured manner based on explicit link with business processes. There are many process mining techniques that can be used in a cost-aware manner; however, it is not clear what impact their limitations will have on the costing system and the research on this topic is relatively scarce. Thus, in our future research, we will analyze the impact of the properties of different process mining techniques on the cost model built on the integration of TDABC and process mining. However, we believe that real-time based cost information can bring significant benefits to organizations through the entire lifecycle of a business process. It empowers them to make informed decisions during the design (e.g., assess feasibility of different process designs and identify potential cost-saving opportunities), implementation (e.g., tracking costs as they occur and compare them against planned budget), monitoring (e.g., track ongoing costs associated with the process), and evaluation phases (e.g., valuable insights for assessing the performance of a business process).

There are various studies in the literature regarding the application of TDABC in different areas, e.g., manufacturing, services, healthcare, logistics, and trading (Gervais et al., 2010; Keel et al., 2017; Rahman et al., 2019; Ribadeneira et al., 2019; Sachini et al., 2022; Santana & Afonso, 2015; Vedernikova et al., 2020; Zamrud & Abu, 2020). Most of the studies on TDABC use time equations with limited focus on the implementation and integration of such costing system or derivation of time equations forming a given costing model. The costing model, time equations and underlying process map are usually derived as a black box based on analysis of underlying activities of the investigated process. The research on the integration of TDABC with process mining, business process simulations, or RPA is scarce. Vom Brocke et al. (2011) examined the intersection of accounting and process-aware information systems and provided a generalized information model based on the ARIS information model and REA accounting model. Later, Sonnenberg and vom Brocke (2014) PAM which can be used to structure event records in process-aware information systems. In this research, we formalized an event log for the purpose of integration of TDABC and process mining for simulation and analysis of costing system. The focus was on the use of cost dimension as a major attribute for the potential implementation of RPA. Sachini et al. (2022) propose use process mining approach for human resource cost calculations. Authors calculate percentages of time spent by human resource on activities within a specific time period, which they define as effort. However, this is very simplistic notion of processing times in the event log, as authors do not consider fluctuations in processing times of such activities, multitasking, batching, and other phenomena related to records of processing times in the event log, which might cause such fluctuations. When it comes to the simulation of TDABC, Sánchez and Batista (2020) proposed a probabilistic TDABC costing model. They use probability distributions to represent time consumption in combination with Monte Carlo simulations. However, in their approach, authors do not consider the process perspective of the simulated process. Moreover, activities are modeled through triangular probability distributions, which is rarely a case in real-life setting. Rahman et al. (2019) also proposed an integration of TDABC and simulation; however, methodology is not explicit when it comes to design and creation of simulation model since time equations were used. To the best of our knowledge, there is no research on the assessment of RPA implementation using the cost dimension of the business process. Our approach could be potentially used in combination with other technologies.

## Conclusions

The focus of this research is on the integration of TDABC and process mining techniques. The objective of the paper was to use the cost dimension as a main attribute for the potential implementation of robotic process automation (RPA). We demonstrated our generic approach toward the implementation of TDABC. The case study uses a real-life event log containing transactional data representing the loan application process. The following procedure was applied to demonstrate our generic approach and provide an answer to *RQ2*. First, the 2011 log was prepared for analysis and simulation purposes. At this stage, it was checked whether all events in the logs contained the basic required attributes in the appropriate formats. Events and related cases that did not possess required attributes or did not respect necessary formats were modified to respect them. Otherwise, they were excluded. Missing values were handled in a similar way. Second, we identified activities involved in the process and discovered the process model of the loan application process contained within the 2011 log. For process discovery, we used split miner, which performs very well among other process discovery techniques and produces process model in BPMN notation. Then, we analyzed the retrieved model to obtain the required process perspectives, e.g., process flow organizational perspective, and resource perspective. Third, based on the analysis, a simulation model representing the original loan application process was designed. The simulation model was checked based on similarity indicators, e.g., process flow and semantics, throughput times, and bottlenecks. Finally, a designed simulation model was then used for the simulation of the partial and full implementation of RPA through separate scenarios (Tab. 3). To answer *RQ1*, we added a cost dimension into the analysis and enriched the event log with cost data based on formalized cost model presented in Section 3. Furthermore, we focused on the data quality when using exact times in the cost model, which is crucial for integrating TDABC and process mining.

Based on analysis of different scenarios of partial and full RPA deployment, we show that even though partial implementation of RPA might not deliver a significant increase in efficiency in the process, it still might deliver significant potential time savings. Tab. 3 shows the impact of the deployment of RPA within the simulation model with respect to individual scenarios. Full automation impacts the average case duration significantly in all scenarios except for activity “W_Assess potential fraud.” The largest changes were achieved in the activities “W_Call after offers,” “W_Validate application” and “W_Handle leads,” namely by 51.4737%, 38.2296%, and 3.8625%, respectively. In terms of reduction of workload, the implementation of RPA achieves the best results for the activities “W_Handle leads,” "W_Validate application," and "W_Call incomplete files." The impact of partial automation on the performance of the discovered process through the deployment of RPA. Partial automation is specified by eliminating the processing times of automated activity. Partial automation significantly impacts average case duration in only two scenarios: "W_Complete application" and "W_Handle leads." Moreover, the reduction of the average duration of the case is less than 0.7590% in all scenarios in the case of partial automation. However, due to the reduction of workload in the case of full automation and partial automation as well, one can see significant potential cost savings in both cases, even though partial automation offers much less savings with regard to efficiency dimension. Moreover, we demonstrated that utilizing exact times from information systems, as opposed to relying on time equations, can substantially disrupt cost allocation. This disruption often arises from employees engaging in multitasking activities (Tabs. 4–5). Consequently, this can undermine the effectiveness of implementing RPA, particularly in cases where multitasking is prevalent within specific tasks.

***Acknowledgements:*** *This paper was supported by the Ministry of Education, Youth and Sports of the Czech Republic within the Institutional Support for Long-term Development of a Research Organisation in 2023.*

## References

Augusto, A., Conforti, R., Dumas, M., Rosa, M., & Bruno, G. (2018). Automated discovery of structured process models from event logs: The discover-and-structure approach. *Data & Knowledge Engineering*, *117*, 373–392. https://doi.org/10.1016/j.datak.2018.04.007

Buijs, J. C. A. M., van Dongen, B. F., & van der Aalst, W. M. P. (2012). On the role of fitness, precision, generalization and simplicity in process discovery. In R. Meersman, H. Panetto, T. Dillon, S. Rinderle-Ma, P. Dadam, X. Zhou, S. Pearson, A. Ferscha, S. Bergamaschi, & I. F. Cruz (Eds.), *On the Move to Meaningful Internet Systems: OTM 2012, Lecture Notes in Computer Science* (Vol. 7565, pp. 305–322). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33606-5_19

Cidav, Z., Mandell, D., Pyne, J., Beidas, R., Curran, G., & Marcus, S. (2020). A pragmatic method for costing implementation strategies using time-driven activity-based costing. *Implementation Science*, *15*(1), 28. https://doi.org/10.1186/s13012-020-00993-1

Garcia, C. dos S., Meincheim, A., Faria Junior, E. R., Dallagassa, M. R., Sato, D. M. V., Carvalho, D. R., Santos, E. A. P., & Scalabrin, E. E. (2019). Process mining techniques and applications – A systematic mapping study. *Expert Systems with Applications*, *133*, 260–295. https://doi.org/10.1016/j.eswa.2019.05.003

Gervais, M., Levant, Y., & Ducrocq, C. (2010). Time-driven activity-based costing (TD-ABC): An initial appraisal through a longitudinal case study. *Journal of Applied Management Accounting Research*, *8*(2), 1–20.

Ghani, N. F. A., Zaini, S. N. A. M., & Abu, M. Y. (2020). Assessment the unused capacity using time driven activity based costing in automotive manufacturing industry. *Journal of Modern Manufacturing Systems and Technology*, *4*(1), 82–94. https://doi.org/10.15282/jmmst.v4i1.3839

Hadid, W., & Hamdan, M. (2022). Firm size and cost system sophistication: The role of firm age. *The British Accounting Review*, *54*(2), 101037. https://doi.org/10.1016/j.bar.2021.101037

Halaška, M. (2021). The importance of lifecycle extension model for implementation of RPA. In J. D. Šebestová, R. Šperka, P. Suchánek, Š. Čemerková, Ž. Rylková, K. Matušínská, R. Bauerová, J. Mazurek, & R. Dolák (Eds.), *Proceedings of 3^rd^ International Conference on Decision Making for Small and Medium-Sized Enterprises* (pp. 173–182). Silesian University in Opava, School of Business Administration in Karviná.

Keel, G., Savage, C., Rafiq, M., & Mazzocato, P. (2017). Time-driven activity-based costing in health care: A systematic review of the literature. *Health Policy*, *121*(7), 755–763. https://doi.org/10.1016/j.healthpol.2017.04.013

Lu, T.-Y., Wang, S.-L., Wu, M.-F., & Cheng, F.-T. (2017). Competitive price strategy with activity-based costing – Case study of bicycle part company. *Procedia CIRP*, *63*, 14–20. https://doi.org/10.1016/j.procir.2017.03.102

Măruşter, L., & Beest, N. R. T. P. (2009). Redesigning business processes: A methodology based on simulation and process mining techniques. *Knowledge and Information Systems*, *21*(3), 267–297. https://doi.org/10.1007/s10115-009-0224-0

Nakatumba, J., Westergaard, M., & Aalst, W. M. P. (2012). Generating event logs with workload-dependent speeds from simulation models. In M. Bajec & J. Eder (Eds.), *Advanced Information Systems Engineering Workshops. Lecture Notes in Business Information Processing* (Vol. 112, pp. 383-397). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31069-0_31

Pashkevich, N., von Schéele, F., & Haftor, D. M. (2023). Accounting for cognitive time in activity-based costing: A technology for the management of digital economy. *Technological Forecasting and Social Change*, *186*, 122176. https://doi.org/10.1016/j.techfore.2022.122176

Popesko, B. (2010). Activity-based costing application methodology for manufacturing industries. *Economics and Management*, *1*(1), 103–114.

Popesko, B., & Tučková, Z. (2012). Utilization of process oriented costing systems in healthcare organizations. *International Journal of Mathematical Models and Methods in Applied Sciences*, *6*(1), 200–208.

Rahman, M. S. bin A., Mohamad, E. bin, & Rahman, A. A. A. (2019). Enhancement of time-driven activity-based costing (TDABC) by using simulation in manufacturing process towards Industry 4.0. *International Journal of Innovative Technology and Exploring Engineering*, *8*(10), 1895–1900. https://doi.org/10.35940/ijitee.j9243.0881019

Ribadeneira, C., Tuapante, L., Siguenza-Guzman, L., Ayabaca, F., Tello, A., & Vanegas, P. (2019). A process collection methodology towards TDABC costing optimization of IT services: A case study in an Ecuadorian university. *Iberian Journal of Information Systems and Technologies*, *1*(E20), 541–552.

Rozinat, A., Mans, R. S., Song, M., & van der Aalst, W. M. P. (2009). Discovering simulation models. *Information Systems*, *34*(3), 305–327. https://doi.org/10.1016/j.is.2008.09.002

Sachini, E., Nikou, E., & Glykas, M. (2022). Process mining human resources cost calculations in time driven activity based costing. In E. N. Degirmenci (Ed.), *New Frontiers for Management and Strategy in the Post-Pandemic Era (Vol. 130)*, *European Proceedings of Social and Behavioural Sciences* (pp. 170–193). European Publisher. https://doi.org/10.15405/epsbs.2022.12.02.15

Sánchez, M. A., & Batista, M. D. (2020). Probabilistic time-driven activity-based costing. *Journal of Corporate Accounting & Finance*, *31*(4), 73–81. https://doi.org/10.1002/jcaf.22468

Santana, A., & Afonso, P. (2015). Analysis of studies on time-driven activity based costing (TDABC). *The International Journal of Management Science and Information Technology*, *15*(1), 133–157.

Sonnenberg, C., & Brocke, J. (2014). The missing link between BPM and accounting: Using event data for accounting in process-oriented organizations. *Business Process Management Journal*, *20*(2), 213–246. https://doi.org/10.1108/bpmj-12-2012-0136

Šperka, R., & Halaška, M. (2023). The performance assessment framework (PPAFR) for RPA implementation in a loan application process using process mining. *Information Systems and E-Business Management*, *21*(2), 277–321. https://doi.org/10.1007/s10257-022-00602-2

Syed, R., Suriadi, S., Adams, M., Bandara, W., Leemans, S. J. J., Ouyang, C., ter Hofstede, A. H. M., van de Weerd, I., Wynn, M. T., & Reijers, H. A. (2020). Robotic process automation: Contemporary themes and challenges. *Computers in Industry*, *115*, 103162. https://doi.org/10.1016/j.compind.2019.103162

Van Dongen, B. (2012). *BPI Challenge 2012* [Dataset]. 4TU.ResearchData. https://data.4tu.nl/articles/dataset/BPI_Challenge_2012/12689204

Vedernikova, O., Siguenza-Guzman, L., Pesantez, J., & Arcentales-Carrion, R. (2020). Time-driven activity-based costing in the assembly industry. *Australasian Business, Accounting & Finance Journal*, *14*(4), 3–23. https://doi.org/10.14453/aabfj.v14i4.2

Vom Brocke, J., Sonnenberg, C., & Baumoel, U. (2011). Linking accounting and process-aware information systems – Towards a generalized information model for process-oriented accounting. *Proceedings of 19th European Conference on Information Systems. ECIS 2011*. https://aisel.aisnet.org/ecis2011/23

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). Engineering psychology and human performance. In *Engineering psychology and human performance*. Psychology Press. https://doi.org/10.4324/9781315665177

Zamrud, N. F., & Abu, M. Y. (2020). Comparative study: Activity based costing and time driven activity based costing in electronic industry. *Journal of Modern Manufacturing Systems and Technology*, *4*(1), 68–81. https://doi.org/10.15282/jmmst.v4i1.3840