UHASSELT
KNOWLEDGE IN ACTION

Faculty of Business Economics

Master of Management

Master's thesis

Visualizing Conformance Checking Results

ADRIJA BOSE

Thesis presented in fulfillment of the requirements for the degree of Master of Management, specialization Data Science

SUPERVISOR :
dr. Gert JANSSENSWILLEN

UHASSELT
KNOWLEDGE IN ACTION
www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

2023
2024

Faculty of Business Economics

Master of Management

Master's thesis

Visualizing Conformance Checking Results

ADRIJA BOSE

Thesis presented in fulfillment of the requirements for the degree of Master of Management, specialization Data Science

SUPERVISOR :

dr. Gert JANSSENSWILLEN

ACKNOWLEDGMENT

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Gert Janssenswillen, for his unwavering guidance, insightful feedback, and continuous support throughout the duration of my master's program. Your expertise and encouragement have been invaluable to the completion of this thesis.

I am grateful to my sister, Ms. Namrata Bose, and my partner for their constant encouragement, patience, and belief in my abilities.

I would also like to extend my heartfelt thanks to my friends, Ms. Angana Dasgupta, Ms. Petra Yaacoub, and Ms. Anwesha Pattnaik, who have provided me with moral support and numerous moments of laughter.

Lastly, I wish to remember my mother, Mrs. Amita Bose, who is no longer with us. Your love, sacrifices, and lessons continue to inspire and guide me every day.

Adrija Bose
June, 2024.

Visualizing Conformance Checking Results

Adrija Bose
Faculty of Business Economics
Hasselt University
Diepenbeek, 3590, Belgium

Abstract— Process mining is widely used in today's data-driven world to comprehensively support understanding and improving business processes. Conformance checking is essential to process mining as it identifies discrepancies between actual process executions and predicted process models. It finds differences between the behavior of the process instances as they actually are (called "as-is") and as they are modelled (called "to-be"). Results indicate that conformance checking is crucial for identifying business process inconsistencies. However, data dependency abstraction and inaccurate event timestamps make it difficult to record and express them. These results show a need for a better and more rigorous conformance checking visualisation. Further work should continue to enhance conformance visualisation tools and incorporate real-time monitoring capabilities in the diagnostics. However, relatively little research has been conducted for effective representations that may present these discrepancies in a simple, intuitive way.

***Keywords*—** Process mining; conformance checking; event logs; visualization techniques

1. Introduction

In current business environments, operating efficiency and compliance are essential. Deviations from intended business processes can arise from system errors, mismanagement of resources, or non-compliance, including deliberate fraud. Organizations must identify and comprehend these discrepancies to effectively manage risks and make well-informed decisions. Process mining has shown to be a useful tool for extracting insights, identifying bottlenecks, and enhancing process efficiency. It does this by using event data from information systems. (Mehr, A. S. M. 2024).

Conformance checking is a rapidly growing subfield of process mining in research and practice. It examines variations in process behaviour by comparing the behaviour of process instances found in an event log to a process model. Conformance checking, thus, quantifies how much process executions deviate from normative or descriptive behaviour. This way, business process managers can evaluate whether their processes work the way they are supposed to, or at least the way they have been depicted in a process model (Dunzer et al., 2019).

Delays, higher costs, lower efficiency, and process interruptions can come from business process deviations. Service quality, customer satisfaction, regulatory compliance, and the company's reputation might also suffer, leading to legal action. They hinder innovation, adaptation, and process improvement. Conformance checking methods find these abnormalities, but they prioritise control flow over data dependencies and resource allocations, misleading diagnostics. Advanced visualisation can improve conformance by checking interpretation, decision-making, and process improvements. Understanding deviation causes and developing conformance checking methods improves process performance and organisational efficiency. This research aims to address the root cause of deviations and different conformance-checking techniques that are available currently, as well as the challenges these techniques pose and looks into the development of advanced visualization techniques to improve the interpretation and communication of the results (Burattin et al., 2018) (Carmona et al., 2018).

Section 2 of this article discusses the important terms and definitions that are required for this paper. Section 3 talks about the problem statement, which includes the research questions. Section 4 covers the methodology and how we answer those questions. Sections 5, 6, 7, and 8 delve into answering the research questions in detail. Section 9 concludes the paper by highlighting the most important points of this paper also while indicating the focus of future research areas.

2. BACKGROUND

2.1. Process Mining

Process mining is a field that focuses on the study of business processes using event logs. It sits at the meeting point of data science and process management. Through knowledge extraction from information systems' event logs, genuine processes can be found, tracked, and improved (van der Aalst, 2011). It involves three main types of analysis: discovery, conformance, and enhancement.

Process discovery is the initial stage of process mining, during which the actual process model is found by applying algorithms to event logs. According to van der Aalst (2016), this model shows a process' actual order of events. Process modelling is putting processes into a formal model so they may be examined or contrasted with event logs to verify compliance. Understanding and enhancing business processes depend on this stage (Dumas et al., 2018). Process enhancement is the use of knowledge from process discovery and conformance checking to enhance current processes. According to van der Aalst (2016), enhancements might entail changing the process model to improve compliance, efficiency, or alignment with corporate objectives.

With an example of the order fulfilment process, let us go through each of the types of process mining, as shown in Figure 1. In process mining, event data is used to find a process model and verify its validity (Leemans et al., 2016). Initially, the company decided to analyse its order fulfilment process using a process discovery technique through which it examined the event logs of past orders. Conformance can be used to evaluate if process documentation is updated or if a discovery model accurately represents the process (Munoz-Gama, 2014). The company then needs to perform conformance checking to ensure process execution matches process discovery. Inefficiencies in their order

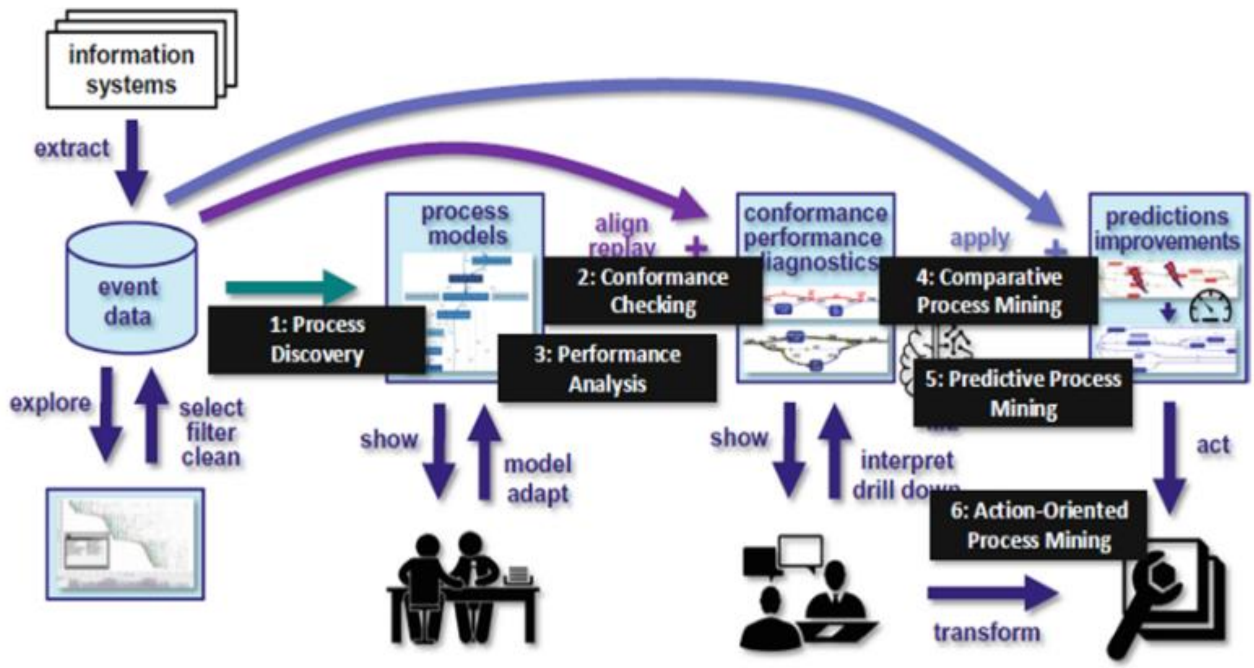

Figure 1: Six frequently used types of process mining by Van der Aalst & Carmona, 2022 p. 24

fulfilment process are identified in this phase. Process mining is used for performance analysis to improve operations (Aalst, 2013). The company can evaluate order fulfilment efficiency and efficacy by comparing its process model to real-world data. Comparative process mining refers to the application of approaches that analyse and compare behaviour, either in the form of models or event logs (van der Aalst, 2014). As part of continuous development, they compare their order fulfilment process to industry best practices or successful competitors to find areas for improvement. Predictive process mining employs single process instances rather than process models to forecast their future (Neu et al., 2022). The company uses predictive process mining to detect order fulfilment bottlenecks. They can forecast future events and streamline processes by examining past data and patterns. The goal of action-oriented process mining is to establish a link between actions and the knowledge gained from event data (Park & van der Aalst, 2022). Finally, using process mining data, the organization optimizes order fulfilment. They employ process discovery, conformance checking, and performance analysis to boost efficiency and customer satisfaction.

2.2. Conformance checking

Conformance checking is the most essential process mining approach, which implies comparing the existing process implementation within the event logs with the given process model. The primary function of conformance checking is to examine the existence or lack thereof of alignment between the stated process behaviour in the model and the actual process run. Inefficiencies may be triggered by analysing discrepancies and deviations in an organization's business processes, leading to compliance with regulations and quality improvements. Conformance checking offers beneficial information about how closely the processes follow their intentional design, enabling organizations to improve their operations, optimize their workflows, and make data-driven decisions. (Carmona et al., 2018)

2.3. Event Logs

Event logs are a collection of events indicating at what point in time, which activity was executed, and for which case. It thus represents the recorded behaviour of a process. Event logs are detailed registers of activities, events, and transactions that happen while the business process is in motion. The log files are records of chronological events that have been documented and are attached to each step of the process, including the sequence of activities, timestamps, and data associated with each event. Event logs are helpful for process mining and analysis, as they can give an overall view of the actual operations of the process in practice. The event logs' ability to record the actions taken within a process allows organizations to track the behaviours of their processes, detect the bottlenecks, detect deviations from the expected behaviour, and analyse the efficiency and effectiveness of their workflows. Event histories are essential in conformance checking, which is a possibility to compare the recorded behaviour with the process model, assess the alignment, find the possible opportunities for improvement and strive for operational excellence. (Carmona et al., 2018)

2.4. Dimensions of Conformance checking

The four most widely recognized dimensions of conformance checking in process mining are fitness, precision, generalization, and simplicity. Fitness evaluates how well a process model matches event log behaviour. A high fitness score indicates the model matches observed behaviour. Precision prevents the model from allowing too much unobserved behaviour. A high precision rating means the model has no incorrect behaviour. The model's capacity to generalize to future behaviour is assessed by generalization. A good generalization ability lets the model handle unknown scenarios, and simplicity aims to keep the model straightforward and easy to understand (Zelst et al., 2017). Organizations can improve their operational effectiveness, strengthen their process models, and make sure their processes follow the original design by taking these factors into account.

3. PROBLEM STATEMENT

Deviations in business processes can have a wide range of effects that significantly influence the efficiency and outcomes of an organisation. These effects can manifest as delays in the execution of processes, escalated operational expenses, diminished productivity, and disruptions in the smooth flow of processes, potentially leading to bottlenecks and inefficiencies. Additionally, such deviations may compromise the quality of service, customer satisfaction, and adherence to regulatory standards, which could culminate in customer discontent, harm to the organisation's reputation, and legal consequences. Furthermore, deviations pose challenges to process improvement initiatives, impede the organisation's ability to adapt, and restrict the potential for innovation and maintaining a competitive edge (Carmona et al., 2018). Additionally, deviations can highlight areas that require improvement to ensure processes are carried out as intended (Burattin et al., 2018). Conformance checking techniques identify and diagnose discrepancies between the behaviour that is seen and that which is represented. Existing conformance checking approaches tend to abstract away data dependencies, resource allocations, and time limitations and instead concentrate on the control flow inside a process. Even when alternative viewpoints are examined, the control flow is aligned first, indicating that one perspective is prioritized. Data dependencies, resource allocations, and time restrictions are only treated as "second-class citizens," which may result in inaccurate conformance diagnostics (Mannhardt et al., 2016). Moreover, existing conformance checking approaches rely significantly on the whole ordering of events; their diagnostics are unreliable and frequently misleading when event timestamps are imprecise or inaccurate (Lu et al., 2015). Additionally, according to Müller et al. (2013), existing techniques may produce false negatives, as they measure conformance based on recorded behaviour without considering potential deviations that may still align with the process specifications.

With these existing problems, we come to the research questions as follows:

RQ 1: Why do deviations happen?

Despite considerable investments in Business Process Management (BPM), deviations still occur, indicating a gap in understanding the factors contributing to these deviations (König et al., 2018). Thus, business process deviations must be analysed to determine their root causes and, therefore, to improve their process efficiency and effectiveness.

RQ 2: What are the different techniques of conformance checking?

Current conformance-checking methods often prioritize control flow over data dependencies, resource assignments, and time constraints, potentially leading to inaccurate diagnostics (Mannhardt et al., 2016). Exploring conformance checking methods is necessary to find the best ones for confirming work process alignment with business process models.

RQ 3: What are the main shortcomings of existing conformance checking techniques?

By identifying and addressing the shortcomings of existing techniques, organizations can enhance the accuracy and reliability of their conformance diagnostics (Mannhardt et al., 2016). Understanding the weaknesses of current methods is essential for developing more robust conformance-checking strategies.

RQ 4: How can a novel visualization technique improve understanding of the conformance checking results?

Novel visualization methods can improve conformance checking interpretation. Advanced visualization displays process variances intuitively, improving decision-making and process improvement.

4. Methodology

A comprehensive literature review of conformance checking methods was carried out in order to identify the shortcomings of the procedures that are currently in place. This was done in order to fill the

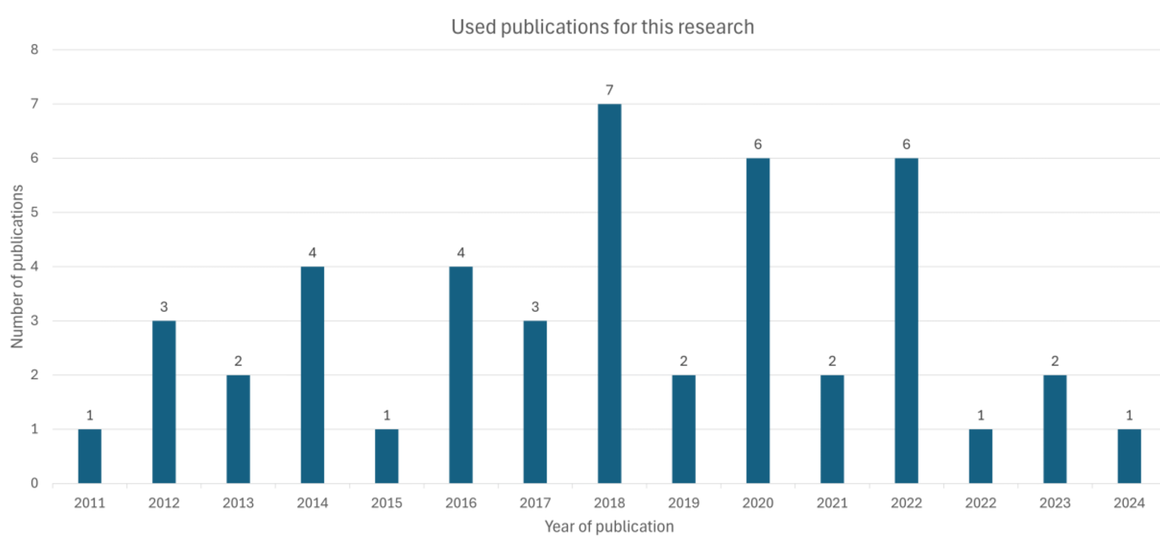

Figure 2 Used publication for this article

research gap that exists regarding effective visualizations that are able to depict and express deviations with more precision. For the purpose of conducting this review, the search phrases “process mining,” “conformance checking," “deviation visualization,” "conformance checking techniques," and "conformance checking drawbacks" were utilized. This article focuses on conformance checking methodologies and provides a review of 45 publications found on Google Scholar. During the process of preparing this paper, we did not take into consideration any papers that were not written in English.

With dates ranging from 2011 to 2024, Figure 2 provides a visual representation of the number of publications that were utilized in the writing of this paper.

5. WHY DO DEVIATIONS HAPPEN?

According to Adriansyah et al., 2011, deviations in conformance checking can happen because of multiple factors. While classical conformance checking techniques are effective in identifying deviations in process executions from predefined models, they can sometimes yield inaccurate results due to strong assumptions. An execution can deviate from the process model to be deemed "acceptable" when an exception not included in the model needs to be handled. For example, a deviation may also be very important if a legal regulation is broken. Examining deviation traces more closely might assist in determining issues related to the process model's quality and how a company's processes are carried out.

The operations being executed in various departments of large corporations are recorded and stored regularly in the IT systems. Using this data from a company's backend, one can apply process mining techniques (such as process discovery) to create a process model. The process model is created using the information coming from the event logs; sometimes, the execution of operations may deviate from the path that was modelled. Deviations can happen due to discrepancies between the expected behaviour defined in the process or ideal behaviour and the actual execution of the process captured in the event log. It can happen because of a lack of coordination between different departments or involved persons, incorrect recording of activity executions, corruption in the recorded event data, and decisions that violate a company's rules. These factors may not be mutually exclusive, and a combination of them can lead to significant consequences for an organization (Adriansyah et al., 2011).

To understand the discrepancies between model predictions and actual observed behaviour, it's useful to categorize the misalignments that are not represented in the model into potential explanations. There are primarily two explanations to consider. Firstly, it could be posited that for any missing event, despite the model indicating a task should have been carried out, the task was indeed performed in that specific process instance. This situation would imply a failure to capture the execution of the task, resulting in its absence from the event log. Alternatively, it might be that the task was genuinely not executed in that instance. This approach can similarly be applied to instances where there is additional behaviour observed in the event log that is not accounted for in the model. It could be inferred that an unmodeled activity was performed or that the recorded event is a false positive. This latter scenario might arise in cases where event logging is manually done or relies on unreliable sensor data, leading to inaccuracies (Carmona et al., 2018).

In order to address deviations found during conformance checking, root cause analysis, or RCA, is essential. It is a systematic approach to investigating and identifying the primary reasons behind problems or discrepancies, aiming to uncover the root cause of deviations. Through detailed RCA, organizations can identify the underlying causes of deviations and implement measures to prevent them from happening again. RCA offers valuable insights into the reasons behind deviations, shedding light on possible vulnerabilities, bottlenecks, or inefficiencies within processes. By understanding and addressing the root causes of deviations, organizations can apply focused interventions to resolve these issues, thereby enhancing their processes' quality, efficiency, and robustness, ultimately boosting their overall performance (Gharahbagheri et al., 2017).

Conducting a comprehensive root cause analysis is pivotal in mitigating these adverse outcomes by uncovering the fundamental causes of deviations. Through detailed root cause analysis, organizations can identify the elements that contribute to discrepancies and implement corrective measures to avert their future occurrence. Employing methods like decision point analysis, rule-based analysis, and decomposed conformance checking can aid in abstracting event data, defining relationships between events, and determining the best alignments to comprehend deviations more accurately. Additionally, incorporating data attributes beyond the control flow, enriching process models with time-related data, and analysing event logs from process-aware information systems can offer deeper insights for effective root cause analysis. Root cause analysis is indispensable for reducing the impact of process deviations, allowing organizations to tackle the root issues efficiently. By adopting sophisticated conformance checking techniques and leveraging data-centric strategies, organizations can improve their capabilities to detect, analyse, and rectify deviations, thereby enhancing process efficiency, ensuring compliance, and boosting overall organisational performance (Carmona et al., 2018).

6. DIFFERENT TECHNIQUES OF CONFORMANCE CHECKING

Conformance checking, a critical part of process mining, emerged as a result of the need to ensure that business processes adhere to predetermined models and their increasing complexity. It compares how processes are being carried out to expected behaviours, allowing organizations to identify deviations, inefficiencies, and non-compliance (Van der Aalst et al., 2012). Conformance checking techniques help analyse the quality of a process model discovered from event data, identify potential deviations, and project real traces onto process models (Carmona & Weidlich, 2022). These techniques evaluate the quality of discovered process models and diagnose deviations from some normative models, thus providing a base for effective visualizations that can represent and communicate deviations more precisely (Leoni et al., 2014).

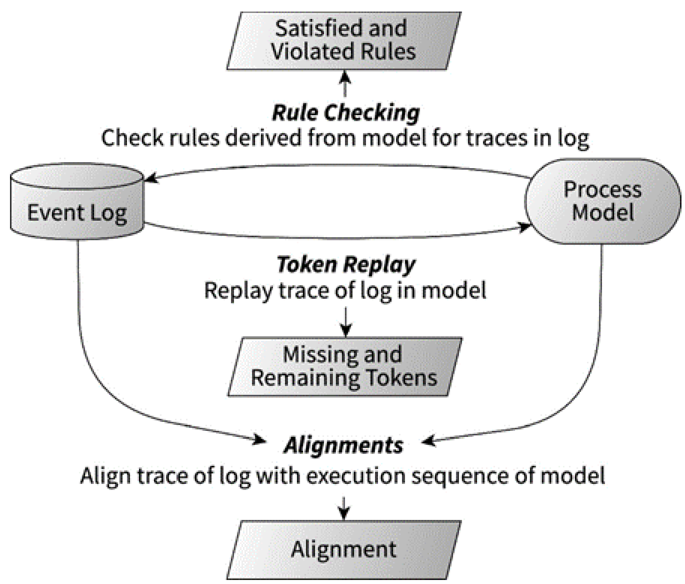

Figure 3: Three different types of conformance-checking, Carmona & Weidlich, 2022 P. 159

Additionally, they cleared the path by introducing the idea of process conformance and highlighting the need to compare process executions with formal process models. The foundation for the advancement of conformance checking methods was thus established. A process model and an event log are required as inputs for compliance verification. It should be made evident when the log and model contradict each other. While there are other approaches to verify compliance, token-based replay and alignments are the two most often used ones (Van der Aalst et al., 2012). The significance of conformance checking in guaranteeing the authenticity of financial information is highlighted by the work of Rozinat et al. (2008), which shows how it may enhance auditing procedures by bringing up anomalies between expected and actual process behaviours.

Existing conformance checking techniques can be classified into the following types: rule checking, rule completeness checking, trace alignment, and stochastic-awareness conformance checking. Rule checking checks to see if the observed behaviour follows the rules provided in the process model, whereas rule completeness checking determines if all of the rules defined in the model are visible in the event log. Trace alignment focuses on aligning individual process traces with the process model and assessing their conformance.

6.1 Rule Checking

Rule checking in conformance checking is a crucial component that includes determining if process instances follow preset rules and restrictions when compared to a particular process model. This verification method checks that process executions match the intended behaviour indicated in the reference model. Organizations may improve reliability, quality, and adherence to best practices by applying specific rules that determine the level of conformance of their processes, systems, or designs with established standards (Dunzer et al., 2019). According to the research by Burattin et al., 2018, using rule-based verification procedures is crucial for verifying that systems comply with their intended designs and processes align with expected behaviours. This emphasis on rule checking highlights its importance in ensuring that operational activities and system executions adhere to the established standards for optimal performance and compliance.

6.2 Token-Based Replay

Token-based replay in conformance checking is a method that involves measuring the number of remaining and missing tokens in a process model when replaying a log. This approach provides insights into the quality of the model and identifies deviations in the log, thereby aiding in diagnosing discrepancies between the expected behaviour and the actual execution of processes (Lu et al., 2015). When replaying the log on the model, the token-based replay approach introduced by Rozinat and van der Aalst counts the number of missing tokens and the remaining tokens and treats the anomalies as deviations.

6.3 Trace Alignment

In conformance checking, trace alignment refers to determining which model trace most closely fits an observed trace of activities from a process instance compared to a reference process model. The purpose of this comparison is to quantify the differences between the behaviour that was seen and the behaviour that was predicted by the process model. Organisations can identify instances where process executions stray from the planned route, assess compliance levels, and find discrepancies using this alignment process. Organisations may increase process efficiency and guarantee adherence to set standards by evaluating the consistency and compliance of operational actions with predetermined process models, which is made possible by the idea of trace alignment. Process performance and compliance may be targeted for improvement through the systematic detection of deviations made possible by this alignment-based analysis (Burattin et al., 2016).

6.4 Stochastic-Aware Conformance checking

Stochastic-aware conformance checking is a technique that evaluates how well process executions match a stochastic process model that considers data dependencies. Unlike traditional conformance checking, which focuses on verifying the compliance of individual traces with a process model, stochastic-aware conformance checking considers the overall distribution of traces in an event log compared to the probability distribution of model executions. Stochastic conformance checking techniques include the Earth Movers' Stochastic Conformance, Entropic Relevance, and Probabilistic Trace Alignments (Mannhardt et al., 2023). Leemans and Polyvyanyy, 2020 emphasize the significance of considering the stochastic nature of the elements being compared and present an entropy-based approach to stochastic conformance verification. The authors point out that the Stochastic Conformance verification technique developed by Earth Movers is an important addition to this subject.

6.4.1 Entropy-based stochastic conformance checking

Entropy-based stochastic conformance checking is a method that utilizes entropy measures to evaluate the alignment between event logs and stochastic process models in process mining. Entropy-based measures are the only quantitative conformance checking techniques that are known to satisfy all the properties for precision and recall that have been proposed till now, including the strict monotonicity properties (Polyvyanyy et al., 2020). Leemans & Polyvyanyy, 2020 in their paper present precision and recall conformance metrics that distinguish between frequent and infrequent deviations between an event log and a process model by quantifying them using the idea of stochastic automaton entropy.

7. What are the main shortcomings of existing conformance checking techniques?

One of the main shortcomings of conformance checking is due to the heavy reliance on the total ordering of events, and their diagnostics are unreliable and often misleading when event timestamps are imprecise or coarse. This limitation makes it harder to accurately identify and communicate the underlying causes of deviations in business operations. It is difficult for the existing conformance checking methods to explain business process irregularities using root-cause analysis (Lu et al., 2015). Additionally, scalability problems present an important risk to the effectiveness of the current conformance checking techniques, particularly when handling large and complicated process models and event logs. Because of processing limitations, the limited scalability of current methods can make it difficult for them to handle the large amounts of data that modern business processes create. This could fail to detect important deviations (Reißner et al., 2019). To overcome the limitations of existing methods, it is imperative to enhance the comprehension and examination of conformance checking findings. Methods from the field of conformance checking allow one to analyse a process model's quality based on event data, spot any deviations, and project actual traces onto process models (Carmona & Weidlich, 2022). The interpretation as well as evaluation of conformance checking findings are improved by these approaches, which are useful for detecting deviations and measuring the accuracy of the discovered model (Sani et al., 2019).

According to Mannhardt et al. (2015), existing conformance checking techniques often focus primarily on control-flow aspects, neglecting data dependencies, resource assignments, and process time constraints. This limited focus can result in incomplete assessments of deviations, as these additional factors are essential for process execution. The offline nature of many conformance checking techniques is another significant drawback. This constraint restricts their utility to fully executed processes and underscores their focus solely on compliance within a specific control-flow framework. This limitation does not include real-time monitoring and intervention capabilities, which delays the discovery and mitigation of deviations in dynamic business contexts (Wang et al., 2022). Rozinat and van der Aalst (2008) present a thorough review of conformance checking in process mining, emphasizing the drawbacks and difficulties of the current conformance checking approaches and highlighting the necessity for sophisticated strategies to deal with these problems. In addition, a review by Carmona et al. (2018) highlights the difficulties and constraints of existing methods by analysing cutting-edge conformance checking techniques in the context of business process management.

Another limitation is the complexity of defining and maintaining rules, especially in complex processes with numerous interdependencies and variations. This complexity can lead to challenges in rule management and may require significant effort to ensure rule accuracy and relevance. Moreover, rule checking may not provide insights into the reasons behind deviations or offer guidance on process improvement (Carmona et al., 2018). Rule checking can only verify whether each event in the log conforms to the specified rules in the process model. However, if the process model is incomplete or inaccurate, the rule derived from the process model may not be comprehensive enough to capture all possible scenarios. Rules written in natural language can be imprecise and difficult to quantify, which limits the effectiveness of rule checking. Formal languages like Petri nets or BPMN can help overcome this. Rules may conflict with each other or with constraints imposed by event log data, making it difficult to reconcile different sources of information. Resolving such conflicts may require manual intervention.

Another significant issue on the subject of conformance checking is computational feasibility. The conformance metrics must become more computationally efficient as event log sizes increase consistently. The larger the event logs get, storing those logs becomes an issue, and therefore, conformance checking needs to function in an online setting, which comes with its own sets of issues (Jans et al., 2021). One key limitation is the static nature of rules, which may not capture real-world processes' dynamic and evolving nature. Rules are predefined and may not adapt well to changes or exceptions in process execution. Additionally, rule checking relies on explicit rules specified in the process model, potentially missing implicit rules or context-specific variations that are crucial for accurate conformance assessment. Checking large event logs against complex process rules can create challenges regarding computational resources and time. In some cases, rule checking may be computationally infeasible.

One significant limitation is that token replay can involve a combinatorial explosion of possible states, which can make it expensive to perform on large event logs or complex models. This limitation leads to incomplete conformance analysis, as events related to unmodeled activities are ignored, potentially masking deviations from the expected behaviour. Additionally, token replay may encounter challenges in missing activities that do not have corresponding tasks in the model, leading to loss of information and accuracy. Token replay can also not provide root cause analysis (Carmona et al., 2018). Additionally, token-based replay faces scalability, timeliness, and traceability issues, particularly in undetermined models (Broucke et al., 2014). The algorithms used in the replay may lead to overestimating metrics due to the artificial creation of superfluous tokens in the model, impacting the accuracy of the conformance assessment. Furthermore, the method may not fully capture the complexities of process behaviour, especially in scenarios involving non-free-choice constructs, potentially leading to incomplete or inaccurate conformance checking results (Bai et al., 2022).

There are a few limitations of trace alignments. The most common limitation of existing alignment techniques is that they are unable to exploit repetitions in the log (Reißner et al., 2020). According to Carrasquel et al. (2021), the alignment method concentrates entirely on control flow aspects, i.e., if system activities adhere to a casual ordering (a message that describes the causal relationship between a message sent event and a message received event). These limitations regulate the comprehensive evaluation of process executions against the model. Trace alignment can be labour-intensive, involves domain knowledge, and is vulnerable to human bias; it may frequently be unavailable. Because of the computational complexity, it might not be feasible to calculate the optimal trace alignment result for a large number of traces (Bose & van der Aalst, 2012). The cost function used to calculate the deviation between the model and the log is sensitive to the parameter settings, which can affect quality and accuracy.

Stochastic conformance checking, too, has a few limitations. As Bortolussi et al. (2022) pointed out, the difficulty in managing parametric stochastic models is a drawback since it might be computationally intensive and impractical to run Stochastic Model Checking (SMC) for every set of parameter values from scratch. This limitation can hinder the practical application of stochastic conformance checking in scenarios where parametric models are prevalent, impacting the ability to verify system behaviours under varying conditions efficiently. Moreover, the lack of support for loops in some stochastic conformance checking techniques limits the types of processes that can be effectively evaluated. This constraint restricts the versatility and comprehensiveness of stochastic conformance checking methods, potentially leading to incomplete assessments of conformance levels in systems with loop structures (Leemans & Polyvyanyy, 2020).

Some solutions to limitations have been tried out. Determining optimal alignments is computationally expensive, particularly considering the increasing amount and complexity of event logs from practice, which can potentially contain traces of several hundred actions and reach one million events. The incapacity of current alignment methods to take advantage of log repetitions is a common drawback. The authors propose a novel approximate technique that uses pre- and post-processing steps to compress the length of a trace and recomputes the alignment cost while ensuring that the cost result never under-approximates the optimal cost by taking advantage of a particular type of sequential pattern in traces, namely tandem repeats. The suggested compression approach systematically outperforms the baselines by up to an order of magnitude in the presence of traces with repetitions, and the cost over-approximation, when it occurs, is negligible, as demonstrated by an extensive empirical evaluation using 50 real-life model log pairs and against six cutting-edge alignment techniques: Alignments of Large Instances (ALI), Evolutionary Approximate Alignments, Trace Sampling, Automata-based Approach with S-Components, Automated Planning Translation, and String Compression Techniques (Reißner et al., 2020).

Reißner et al. (2020) emphasize their approaches—TR-SComp (S-Components and tandem repeat reduction) and Hybrid (a hybrid approach that tries to automatically select the most suitable extension based on the characteristics of the input model and log). After evaluating 50 real-life model-log pairs, they illustrate that their technique, in the presence of impactful repetitive behaviour in the log, after applying on top of the automata-based approach, systematically outperforms five baseline techniques. This suggested method is relevant to concurrency-free process models. Yet, the authors demonstrate how this technique can be linked in a decomposition framework so that it may be used in models in order to display concurrency. Reißner et al. (2020) suggest using S-Component decomposition (S-Component in a Petri net is a subnet consisting of places and transitions, where each place has exactly one incoming and one outgoing edge) as it automatically yields concurrency-free process models. Although the technique is not limited to this exact decomposition process, alternative decomposition strategies can be applied as long as they result in concurrent process models. Their paper also addressed the issue of identifying unfitting log behaviour.

In another paper, Berti & Aalst (2021) introduce an improved token-based replay approach that aims to accelerate conformance checking processes and is much faster and more scalable. Additionally, the method offers more precise diagnostics that help identify compliance problems and prevent well-known issues like "token flooding." The PM4Py (Python library for process mining) has adopted the unique token-based replay mechanism. The authors explain how conformance checking is taken over by more advanced approaches like alignment, and token-based replay got discarded. Through this paper, Berti & Aalst (2021) reinforce a rejuvenation of token-based replay. The approach improves the execution time of the token-based replay operation, increasing the performance gap between token-based replay and alignments. The approach uses root cause analysis as a diagnostic (on the token-based replay output). However, Berti & Aalst (2021), while exploring the novel technique, also highlight the limitations. They do not suggest any fitness or termination assurances. Also, in certain situations, performance is worse than sophisticated replay methods like automaton-based alignments (as AFA).

8. How can a novel visualization technique improve understanding of the conformance checking results?

Visualisation techniques can significantly improve the understanding of conformance checking results by providing intuitive visualizations of the relationships between observed behaviour and modelled behaviour. These visualizations can aid in diagnosing discrepancies, identifying main paths of execution, and ultimately enhancing the quality and compliance of processes. Kriglstein et al. (2016) analyse how process mining techniques can support visual analytical aspects. They propose a first approach to how such techniques can be categorized by adopting frequently used ProM plug-ins. They conducted their study to analyse how visual analytic aspects are supported. The authors recognize the factor that it was helpful for users who were not very familiar with various process mining techniques. The results of this paper also have shown that the categories were overlapping and not exhaustive. However, they do a good job of characterizing process mining approaches through visual analytics features.

After the above paper, Dixit et al., (2017) explore more by introducing a novel tool for enabling interactive process-oriented data analysis. The tool enables interactive process analysis by utilizing current approaches from the fields of process mining, data mining, and visual analytics. It allows for exploratory analysis to be conducted by providing a variety of helicopter perspectives on the process. In contrast to the methodologies that are now in use, it is very interactive, which means that it can be utilized to carry out root cause analysis for any issues that may arise during the process. An analysis of a real-world dataset was performed with the help of the tool, which has a wide range of application areas that can be classified into broad categories. In order to do process analytics, the tool makes use of the conventional methods of data representation, such as histograms. However, their tool is currently limited to read-only data plots.

In a study by Rehse et al., (2022), commercial tools offer a more sophisticated analysis pipeline for visualizing conformance checking compared to academic tools. Commercial tools provide a systematic analytical procedure that involves measuring the overall adherence to a process, breaking down adherence values across different aspects, identifying specific deviations in process models, and analysing particular deviations with possible underlying reasons. Academic tools mostly concentrate on the first and final stages of this process, as opposed to other stages. The findings highlight the differences between commercial and academic tools in conformance checking visualization, suggest areas for future research to improve visualization techniques, and propose a framework to guide further research in this domain.

In order to promote quality improvement and help practitioners understand local care processes, Dahlin, (2020) considers visualization essential. Visualization tools, such as Lexis diagrams, can drive quality improvement initiatives by offering understandable and intuitive representations of complicated data, hence aiding in the comprehension of local care processes. The author adds incorporating components to diagrams requires caution. Plotting data, for example, individual patient details, requires caution because they can't compensate for case differences. The amount of dots each year may depend on disease prevalence, not just lifespan. Relapses could shorten people's lives or disease duration, making the basic picture problematic. A 3D visualization with another time axis could solve this. The study demonstrated that Lexis diagrams could aid in understanding survival data, trigger important dialogues among care providers, support quality improvement efforts, offer new perspectives, and complement traditional survival curves in the context of gynaecological cancer care quality improvement.

Gschwandtner, (2017) explains how the combination of visual analytics with process mining has the potential to obtain a more profound understanding of process behaviour, enhance processes, and make data-informed decisions to enhance organizational performance. Visual analytics is a useful tool for revealing concealed patterns, connections, and irregularities in event data. It aids in the identification of patterns, verification of compliance, and improvement of processes. Visual approaches, such as flow charts, directed graphs, and event sequence visualizations, are commonly used to represent process models, identify patterns, and check conformance in event data. The integration of visual analytics with process mining provides advantages such as increased data exploration, improved insight generation, efficient pattern recognition, decision-making support, iterative analysis, and scalability in evaluating substantial volumes of event data. The author talks about various challenges that they faced while combining visual analytics and process mining in the paper, like the complexity of data, the need for effective data exploration and understanding, and the evaluation of visualizations in the context of process analysis.

Wunderlich et al. (2017), in a study to explore the delay of trains, demonstrate a novel interactive visualization design. They consider two variants- cumulative and non-cumulative delay distribution in order to display the delay of trains in a better way and thus showcase the potential impact on travel. It provides information regarding the repercussions of the delay on both the timeliness of arrival and the ability to catch connecting trains. Additionally, it presents alternate rail connections in the event of delays. They performed a user study to assess the effectiveness of their design. By comparing their design with two existing displays, they found that their design demonstrates a strong comprehension and has a favourable effect on trip planning by displaying predicted delays and their effects. Conversely, the current displays do not provide important information regarding delays for planning issues that prioritize delay reduction.

9. CONCLUSION

Process mining involves conformance checking, according to this literature study. Conformance checking compares process executions to models to discover inconsistencies and assure standards compliance. Inefficiencies, fraud, and regulatory compliance are detected by conformance checking process executions against modelled expectations. Organisations need this procedure to optimise, reduce risks, and maintain process integrity.

This study evaluates rule checking, token-based replay, trace alignment, and stochastic-aware conformance checking. Though effective, some treatments have drawbacks. Misuse of event total ordering might lead to incorrect diagnosis if event timestamps are poor. Data dependencies, resource allocations, and time limits are regularly ignored, worsening this problem. Process deviation assessments are vague and disguise restrictions-related inefficiencies. Conventional approaches lose efficiency and accuracy as event logs grow. Offline methods often limit real-time monitoring and action.

Researchers suggest future directions where improved conformance checking to circumvent these limits is necessary. Novel visualisation methods for variations and causes are promising and displaying sophisticated conformance checking findings boosts performance. Focus on real-time monitoring can help organisations fix conformance checking difficulties. Diagnostic accuracy and relevance would improve with a holistic perspective of process adherence. AI and ML can scale big dataset conformance checking. Multi-perspective approaches help companies assess and optimise their operations, assuring compliance, lowering risks, and enhancing efficiency. Studying multi-perspective conformance checking approaches in the future that consider control flow, data dependencies, resource allocations, and timing can help. Researchers also suggest that interactive data exploration, like plotted graph views, is advantageous along with data plots. This study implies conformance checking innovation, and research improves process management.

References

[1] Adriansyah, A., Sidorova, N., & van Dongen, B. F. (2011, June). Cost-based fitness in conformance checking. In 2011 eleventh international conference on application of concurrency to system design (pp. 57-66). IEEE.

[2] Bai, E., Qi, M., Luan, W., Li, P., & Du, Y. (2022). Repairing Process Models with Non-Free-Choice Constructs Based on Token Replay. Computing & Informatics, 41(4).

[3] Berti, A., & van der Aalst, W. M. (2021). A novel token-based replay technique to speed up conformance checking and process enhancement. In Transactions on Petri Nets and Other Models of Concurrency XV (pp. 1-26). Berlin, Heidelberg: Springer Berlin Heidelberg.

[4] Bortolussi, L., Cairoli, F., Carbone, G., & Pulcini, P. (2022). Stochastic variational smoothed model checking. arXiv preprint arXiv:2205.05398.

[5] Bose, R. J. C., & van der Aalst, W. M. (2012). Process diagnostics using trace alignment: opportunities, issues, and challenges. Information Systems, 37(2), 117-141.

[6] Burattin, A., Maggi, F. M., & Sperduti, A. (2016). Conformance checking based on multi-perspective declarative process models. Expert systems with applications, 65, 194-211.

[7] Burattin, A., van Zelst, S. J., Armas-Cervantes, A., van Dongen, B. F., & Carmona, J. (2018). Online conformance checking using behavioural patterns. In Business Process Management: 16th International Conference, BPM 2018, Sydney, NSW, Australia, September 9–14, 2018, Proceedings 16 (pp. 250-267). Springer International Publishing.

[8] Carmona, J., van Dongen, B., & Weidlich, M. (2022). Conformance checking: foundations, milestones and challenges. In Process Mining Handbook (pp. 155-190). Cham: Springer International Publishing.

[9] Carmona, J., van Dongen, B., Solti, A., & Weidlich, M. (2018). Conformance checking. Switzerland: Springer.[Google Scholar], 56.

[10] Dahlin, S. (2020). Exploring the usefulness of Lexis diagrams for quality improvement. BMC Medical Informatics and Decision Making, 20(1), 7.

[11] de Leoni, M., Munoz-Gama, J., Carmona, J., & van der Aalst, W. M. (2014). Decomposing alignment-based conformance checking of data-aware process models. In On the Move to Meaningful Internet Systems: OTM 2014 Conferences: Confederated International Conferences: CoopIS, and ODBASE 2014, Amantea, Italy, October 27-31, 2014, Proceedings (pp. 3-20). Springer Berlin Heidelberg.

[12] Dixit, P. M., Caballero, H. G., Corvo, A., Hompes, B. F. A., Buijs, J. C., & van der Aalst, W. M. (2017, February). Enabling interactive process analysis with process mining and visual analytics. In Special Session on Analysis of Clinical Processes (Vol. 6, pp. 573-584). SCITEPRESS.

[13] Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). Fundamentals of Business Process Management. Springer

[14] Dunzer, S., Stierle, M., Matzner, M., & Baier, S. (2019, June). Conformance checking: a state-of-the-art literature review. In Proceedings of the 11th international conference on subject-oriented business process management (pp. 1-10).

[15] Fani Sani, M., van Zelst, S. J., & van der Aalst, W. M. (2020). Conformance checking approximation using subset selection and edit distance. In Advanced Information Systems Engineering: 32nd International Conference, CAiSE 2020, Grenoble, France, June 8–12, 2020, Proceedings 32 (pp. 234-251). Springer International Publishing.

[16] Gschwandtner, T. (2017). Visual analytics meets process mining: challenges and opportunities. In Data-Driven Process Discovery and Analysis: 5th IFIP WG 2.6 International Symposium, SIMPDA 2015, Vienna, Austria, December 9-11, 2015, Revised Selected Papers 5 (pp. 142-154). Springer International Publishing.

[17] Jans, M., De Weerdt, J., Depaire, B., Dumas, M., & Janssenswillen, G. (2021). Conformance checking in process mining. Information Systems, 102, 101851.

[18] Kriglstein, S., Pohl, M., Rinderle-Ma, S., & Stallinger, M. (2016, June). Visual Analytics in Process Mining: Classification of Process Mining Techniques. In EuroVA@ EuroVis (pp. 43-47).

[19] Leemans, S. J., & Polyvyanyy, A. (2020). Stochastic-aware conformance checking: An entropy-based approach. In Advanced Information Systems Engineering: 32nd International Conference, CAiSE 2020, Grenoble, France, June 8–12, 2020, Proceedings 32 (pp. 217-233). Springer International Publishing.

[20] Leemans, S. J., Fahland, D., & Van der Aalst, W. M. (2018). Scalable process discovery and

conformance checking. Software & Systems Modeling, 17, 599-631.

[21] Loyola-González, O. (2023). Process mining: software comparison, trends, and challenges. International Journal of Data Science and Analytics, 15(4), 407-420

[22] Lu, X., Fahland, D., & van der Aalst, W. M. (2015). Conformance checking based on partially ordered event data. In Business Process Management Workshops: BPM 2014 International Workshops, Eindhoven, The Netherlands, September 7-8, 2014, Revised Papers 12 (pp. 75-88). Springer International Publishing.

[23] Mannhardt, F., Bovo, R., Oliveira, M. F., & Julier, S. (2018, November). A taxonomy for combining activity recognition and process discovery in industrial environments. In International Conference on Intelligent Data Engineering and Automated Learning (pp. 84-93). Cham: Springer International Publishing

[24] Mannhardt, F., De Leoni, M., Reijers, H. A., & Van Der Aalst, W. M. (2016). Balanced multi-perspective checking of process conformance. Computing, 98, 407-437.

[25] Mannhardt, F., Leemans, S. J., Schwanen, C. T., & de Leoni, M. (2023, May). Modelling data-aware stochastic processes-discovery and conformance checking. In International Conference on Applications and Theory of Petri Nets and Concurrency (pp. 77-98). Cham: Springer Nature Switzerland.

[26] Mehr, A. S. M. (2024). Multi-perspective Conformance Checking: Identifying and Understanding Patterns of Anomalous Behavior.

[27] Müller, R., van der Aalst, W. M., & Stahl, C. (2013). Conformance checking of services using the best matching private view. In Web Services and Formal Methods: 9th International Workshop, WS-FM 2012, Tallinn, Estonia, September 6-7, 2012, Revised Selected Papers 9 (pp. 49-68). Springer Berlin Heidelberg.

[28] Munoz-Gama, J. (2014). Conformance checking and diagnosis in process mining (Doctoral dissertation, Polytechnic University of Catalonia, Spain).

[29] Neu, Dominic A., Johannes Lahann, and Peter Fettke. "A systematic literature review on state-of-the-art deep learning methods for process prediction." Artificial Intelligence Review 55.2 (2022): 801-827.

[30] Park, G., & van der Aalst, W. M. (2022). Action-oriented process mining: bridging the gap between insights and actions. Progress in Artificial Intelligence, 1-22.

[31] Polyvyanyy, A., Moffat, A., & García-Bañuelos, L. (2020, October). An entropic relevance measure for stochastic conformance checking in process mining. In 2020 2nd International Conference on Process Mining (ICPM) (pp. 97-104). IEEE.

[32] Rehse, J. R., Pufahl, L., Grohs, M., & Klein, L. M. (2022). Process mining meets visual analytics: the case of conformance checking. arXiv preprint arXiv:2209.09712.

[33] Reißner, D., Armas-Cervantes, A., & La Rosa, M. (2020). Efficient Conformance Checking using Approximate Alignment Computation with Tandem Repeats. arXiv preprint arXiv:2004.01781.

[34] Reißner, D., Armas-Cervantes, A., Conforti, R., Dumas, M., Fahland, D., & La Rosa, M. (2020). Scalable alignment of process models and event logs: An approach based on automata and s-components. Information Systems, 94, 101561.

[35] Tian, Y., Du, Y., Li, M., Han, D., & Hu, Q. (2018). Reduced alignment based on Petri nets. Concurrency and Computation: Practice and Experience, 30(23), e4411.

[36] Van der Aalst, W. M. (2013). Decomposing Petri nets for process mining: A generic approach. Distributed and Parallel Databases, 31, 471-507.

[37] van der Aalst, W. M. (2014). How people really (like to) work: comparative process mining to unravel human behavior. In Human-Centered Software Engineering: 5th IFIP WG 13.2 International Conference, HCSE 2014, Paderborn, Germany, September 16-18, 2014. Proceedings 5 (pp. 317-321). Springer Berlin Heidelberg.

[38] Van Der Aalst, W., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., ... & Wynn, M. (2012). Process mining manifesto. In Business Process Management Workshops: BPM 2011 International Workshops, Clermont-Ferrand, France, August 29, 2011, Revised Selected Papers, Part I 9 (pp. 169-194). Springer Berlin Heidelberg

[39] Van Der Aalst, Wil, and Wil van der Aalst. Data science in action. Springer Berlin Heidelberg, 2016

[40] van Zelst, S. J., van Dongen, B. F., & van der Aalst, W. M. (2018). Event stream-based process discovery using abstract representations. Knowledge and Information Systems, 54, 407-435.

[41] vanden Broucke, S. K., Munoz-Gama, J., Carmona, J., Baesens, B., & Vanthienen, J. (2014). Event-based real-time decomposed conformance analysis. In On the Move to Meaningful Internet Systems: OTM 2014 Conferences: Confederated International Conferences: CoopIS, and ODBASE 2014, Amantea, Italy, October 27-31, 2014, Proceedings (pp. 345-363). Springer Berlin Heidelberg.

[42] Wang, J., Chang, V., Yu, D., Liu, C., Ma, X., & Yu, D. (2022). Conformance-oriented predictive process monitoring in BPaaS based on combination of neural networks. Journal of Grid Computing, 20(3), 25.

[43] Wunderlich, M., Ballweg, K., Fuchs, G., & von Landesberger, T. (2017, June). Visualization of delay uncertainty and its impact on train trip planning: A design study. In Computer Graphics Forum (Vol. 36, No. 3, pp. 317-328).

[44] König, U. M., Linhart, A., & Röglinger, M. (2019). Why do business processes deviate? Results from a Delphi study. Business Research, 12, 425-453.

[45] Van der Aalst, W., Adriansyah, A., & Van Dongen, B. (2012). Replaying history on process models for conformance checking and performance analysis. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(2), 182-192.