ProcessProfiler3D

Wynn, M.T.; Poppe, E.; Xu, J.; ter Hofstede, A.H.M.; Brown, R.; Pini, A.; van der Aalst, W.M.P.

Published in:
Decision Support Systems

DOI:
10.1016/j.dss.2017.04.004

Published: 01/08/2017

Decision Support Systems 100 (2017) 93–108

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

ProcessProfiler3D: A visualisation framework for log-based process performance comparison

M.T. Wynn[a,*], E. Poppe[a,*], J. Xu[a], A.H.M. ter Hofstede[a,b], R. Brown[a], A. Pini[c], W.M.P. van der Aalst[b,a]

[a] Queensland University of Technology, Queensland, Australia
[b] Eindhoven University of Technology, Eindhoven, The Netherlands
[c] DensityDesign Research Lab, Politecnico di Milano, Milan, Italy

ARTICLE INFO

Available online 2 May 2017

Keywords:
Process analysis
Process mining
Performance analysis
Visualisation

ABSTRACT

An organisation can significantly improve its performance by observing how their business operations are currently being carried out. A great way to derive evidence-based process improvement insights is to compare the behaviour and performance of processes for different process cohorts by utilising the information recorded in event logs. A process cohort is a coherent group of process instances that has one or more shared characteristics. Such process performance comparisons can highlight positive or negative variations that can be evident in a particular cohort, thus enabling a tailored approach to process improvement. Although existing process mining techniques can be used to calculate various statistics from event logs for performance analysis, most techniques calculate and display the statistics for each cohort separately. Furthermore, the numerical statistics and simple visualisations may not be intuitive enough to allow users to compare the performance of various cohorts efficiently and effectively. We developed a novel visualisation framework for log-based process performance comparison to address these issues. It enables analysts to quickly identify the performance differences between cohorts. The framework supports the selection of cohorts and a three-dimensional visualisation to compare the cohorts using a variety of performance metrics. The approach has been implemented as a set of plug-ins within the open source process mining framework ProM and has been evaluated using two real-life datasets from the insurance domain to assess the usefulness of such a tool. This paper also derives a set of design principles from our approach which provide guidance for the development of new approaches to process cohort performance comparison.

1. Introduction

Performance analysis can provide valuable insights into business processes of organisations, such as where bottlenecks and waiting times occur. Such analyses provide a valuable starting point for business process redesign aimed at cost reduction, time savings and/or productivity gains. Process mining [1], a specialised field of research in business process management, uses process-related data, recorded in event logs to uncover the real behaviour and performance of business processes.

A key limitation of contemporary process mining techniques for performance analysis is the lack of support for a detailed comparison of the characteristics of multiple cohorts of process instances. We consider a process cohort to be *a group of process instances that has one or more shared characteristics*. For example, one may wish to be able to contrast the processing of low-value claims (cohort 1) with that of high-value claims (cohort 2) in an insurance company or to gain insight into the different pathways of patients through an emergency department based on the severity of their injuries, as reported in several industry case studies [2–4]. While the performance statistics for each process cohort in these scenarios were computed using existing process mining techniques, the detailed comparison itself is still a manual process. Only recently, a few approaches have demonstrated the visual comparison of two process cohorts in one analysis. However, the comparison of more than two cohorts in a single visualisation is still not well-developed.

The focus of this paper, therefore, is to present a novel comparative process visualisation framework that supports process performance comparison for multiple cohorts in a single visualisation

* Corresponding authors at: Queensland University of Technology, School of Information Systems, 2 George St, Brisbane QLD 4000, Queensland, Australia.
E-mail addresses: m.wynn@qut.edu.au (M. Wynn), erik.poppe@qut.edu.au (E. Poppe), j15.xu@qut.edu.au (J. Xu), a.terhofstede@qut.edu.au (A.t. Hofstede), r.brown@qut.edu.au (R. Brown), azzurra.pini@polimi.it (A. Pini), w.m.p.v.d.aalst@tue.nl (W. van der Aalst).

http://dx.doi.org/10.1016/j.dss.2017.04.004
0167-9236/

and a set of generalised design principles to guide the development of further approaches for process cohort performance comparison. Fig. 1 shows an overview of the approach. The framework takes an event log and a representative process model as inputs and supports dynamic categorisation of process cohorts and computation of performance metrics. The framework subsequently generates comparative process visualisations, which can then be interactively explored by process stakeholders. These visualisations extend the two-dimensional graphs proposed by Pini et al. [5] to a three-dimensional space, creating a new visual technique that enables global comparison. The rest of the paper is organised as follows. Section 2 discusses related work, while in Section 3 the approach proposed in this paper is discussed in detail. In Section 4 aspects of the implementation of the framework are discussed and Section 5 is concerned with the framework's evaluation. Section 6 summarises our contributions and presents some avenues for future work.

2. Related work

Information systems nowadays are designed to automatically capture process related data, such as activities, time stamps, resources and contextual information (e.g. customer details, case-specific data) [1], which has enabled the application of process mining techniques to obtain indicators of organisational performance from such data. In addition to academic interest in this topic [6–9], there are currently over 20 commercial process mining tools that include performance analysis features [1], mostly comparable to the features provided by Disco [10]. Most of these techniques analyse a single dimension of a business process at a time. However, in practice, process stakeholders are often interested in performance comparisons which focus on comparing performance between process instances with different characteristics (e.g. performance differences between teams, types of applications) [2–4]. Comparing variants of the same process in such a way can help identify root causes of process performance differences. In such scenarios process instances from an event log are partitioned into multiple cohorts based on different attributes (e.g. resource, case attribute, context variables). The behaviour and performance of these processes for different cohorts are then compared. Due to a lack of tool support for such comparative analyses, they are commonly performed by expert process analysts analysing each cohort individually and then manually comparing the results as reported in several recent case studies [2–4]. Suriadi et al. [2] report on comparing cohorts in the event log of a major insurance company. They report having to use a complex chain of tools for the process, dealing repeatedly with filtering procedures to split the event log into meaningful cohorts and encountering interoperability issues between various tools. Consequently, they described the analysis process as "time-consuming, tedious and error-prone". Partington et al. [3] present a comparative study of processes at four hospitals. They use side-by-side comparisons of process maps to identify differences in process flow, but report that there is a lack of tools to support this kind of analysis. While some recent work [4] has focussed on automating the involved log filtering and metric computation, these approaches still present the analysis results for each cohort in separate windows and rely on process analysts to spot the differences. As such, these approaches still make the process both error prone and not very scalable. All these studies required significant effort including massaging the data, using multiple process mining tools and manually comparing multiple visualisations. Consequently, receiving an answer to the simple analysis question "How does my process variant A differ from process variant B?", required organisations in all three cases to engage research teams over extended periods of time.

A research question of interest to both researchers and industry is therefore: "How can the comparison of process cohorts be supported in an intuitive way to enable non-expert stakeholders to gain insights from their process data?". Four high-level requirements can be derived out of the presented analysis scenarios and can be used to assess the suitability of existing approaches for the presented analysis scenarios.

Firstly, the researchers had to filter cohorts to compare out of one or more event logs. This filtering step sometimes had to be repeated iteratively to identify boundaries of some cohorts or as the researchers' understanding of the data evolved. Being able to define cohorts and split event logs interactively, therefore, appears to be an important feature for comparing multiple cohorts. While many tools already support interactive filtering of the event log being analysed, some approaches [6,11-14] expect the user to pre-filter or annotate event logs using separate tools. This increases the danger of interoperability issues and generally makes interactive exploration of the data infeasible. On the other hand, several recent approaches [4,15,16] apply data warehousing techniques to enable users to interactively specify and explore process cohorts by proposing the concept of "process cubes".

Secondly, the researchers needed to calculate metrics to conceptualise process performance for different parts of the analysed process. These metrics all related to frequencies of occurrence and the time perspective of the process [3]. Two issues are of relevance to the computation of these metrics. The first issue in the calculation of process performance from an event log is whether the analysis is aware of the structure of the underlying process or not. For example, many existing approaches do not consider parallelism of activities, which can distort the results of the analysis [1]. The second issue is whether the approach can compute the metrics for the comparison interactively, as otherwise interoperability with other tools or interactivity are again problematic.

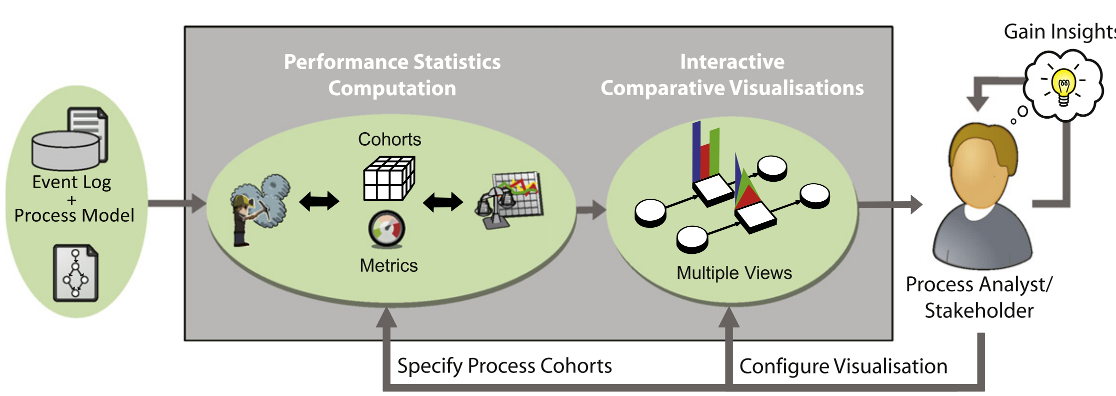


Fig. 1. Overview of the proposed approach for process performance comparisons.

Thirdly, the researchers had to compare these metrics for different parts of the process being analysed. Notably, all existing approaches to cohort performance analysis use visualisation to present and compare performance metrics. For this task, they mostly used side-by-side comparisons of visualised metrics which meant they had to manually compare visual attributes, such as colour and line thickness, across multiple visualisations. Many existing approaches, such as existing process mining [6,7,10] and process cube tools [4,15,16] present their analysis results in such a fashion. However, approaches using juxtaposition, i.e. side-by-side images or even matrices of images, are problematic as such comparisons require the analyst to spot the difference between multiple visualisations. This process is cognitively inefficient and scales badly when more than two values need to be compared [17]. Instead, an integrated approach to present the performance data in one visualisation could effectively offload some of the comparison effort onto the visual perceptual system using preattentive processing [18]. For example, Minit [19], the visual analytics tools for event log comparison [20–22] and recent comparative visualisation approaches [11–13] do provide integrated visualisations for the comparison of two cohorts to facilitate understanding cohort differences. However, most of these approaches visually present the difference between cohorts, rather than the performance indicator values themselves. Consequently, they are limited to comparing two process cohorts at a time. Van Mourik [12] presents the only approach that presents performance metrics for more than two cohorts in one integrated visualisation.

Fourthly, the researchers mentioned interoperability issues between the numerous tools they used for the analysis slowing down the process. An integrated approach should therefore be able to handle all three stages of the comparative analysis: splitting the event log into cohorts, computing the performance metrics for each cohort and presenting the differences between cohorts to the user.

The proposed approach therefore fills a gap as it is the only solution that fully covers analysis scenarios such as presented by Partington et al. [3], by providing both an integrated and interactive approach to multi-cohort process performance comparison. Our previous work [5] proposed projecting performance statistics from multiple cohorts on to a process model in the form of bar and triangle charts. Some designs of visual graphs are adopted from there. However, the visualisations shown in [5] were manually drawn using process statistics computed by another application and a simplified process model. The approach used for visualising waiting times differences between cohorts is also problematic as it is based on a linear layout of the process. To address these issues we revised our approach to use three dimensional visualisation to visualise performance data on top of the process model rather than setting the process model elements into different sizes and colours. With the additional dimension, more metrics and multiple cohorts can be displayed in one integrated view. It also enables users to interactively split an event log and thus to easily explore different cohorts in the dataset.

In summary, comparing the performance of process cohorts is a problem of interest to industry and academia. However, while various approaches have been presented that support different parts of such an analysis, existing work on visual comparison of the performance of multiple process cohorts using event logs still does not provide the full capabilities required for some analysis scenarios of interest to industry, as shown in Table 1.

3. Approach

As the literature review shows there is a lack of tools that enable an intuitive and direct comparison of multiple process cohorts. In order to address this issue, we propose a novel tool that enables a visual comparison of such cohorts. This tool has been designed following a Design Science methodology as discussed in Section 3.1 by identifying requirements for process cohort comparison in Section 3.2 and proposing design principles to satisfy these requirements in Section 3.5. These requirements have been met by computing performance indicators, as described in Section 3.3 and applying visualisation theory, as described in Section 3.4.

Table 1
Comparison of existing tools for comparative process performance analysis.

Tools	Cohort split	Metric computation		Comparison capabilities				Results presentation	
	Interactive cohort exploration	Considers parallelism	Process metric computation	Process frequencies	Process timing	Two cohorts	Many cohorts	Integrated two cohorts	Integrated many cohorts
ProM Replay plugin [6]	**No**	Yes	Yes	Yes	Yes	**No**	**No**	**No**	**No**
ProM Inductive Visual Miner [7], Disco [10]	Yes	Yes	Yes	Yes	Yes	**No**	**No**	**No**	**No**
Minit [19]	Yes	Yes	Yes	Yes	Yes	Yes	**No**	Yes	**No**
ProCube [15], PMC [4]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	**No**	**No**
PMCube [16]	Yes	Yes	Yes	Yes	**No**	Yes	Yes	Yes	**No**
Misue [23]	Yes	**No**	**No**	Yes	Yes	Yes	Yes	**No**	**No**
Buijs [14]	**No**	Yes	Yes	**No**	**No**	Yes	Yes	**No**	**No**
Malik et al. [20]	Yes	**No**	Yes	Yes	Yes	Yes	**No**	Yes	**No**
Basole et al. [21]	Yes	**No**	Yes	Yes	**No**	Yes	**No**	Yes	**No**
Zhang et al. [22]	Yes	**No**	Yes	Yes	Yes	Yes	**No**	Yes	**No**
Bolt et al. [11]	**No**	**No**	Yes	Yes	Yes	Yes	**No**	Yes	**No**
Kriglstein et al. [13]	**No**	Yes	Yes	Yes	**No**	Yes	**No**	Yes	**No**
Van Mourik [12]	**No**	Yes	**No**	Yes	Yes	Yes	Yes	Yes	Yes
ProcessProfiler3D	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Bold entries highlight gaps in current tool support for comparative process cohort performance analysis.

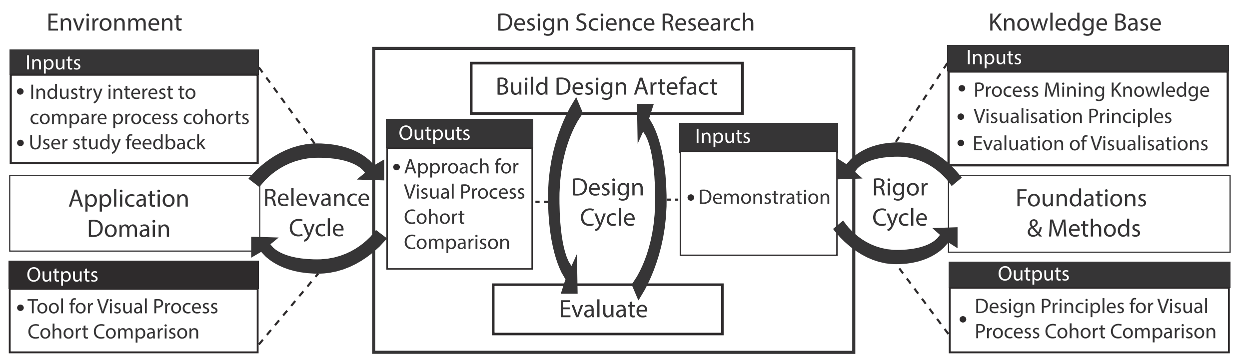


Fig. 2. Methodology overview.
Source: Adapted from Hevner [25].

3.1. Methodology

The goal of the presented work is to develop a tool that enables a detailed comparison of the performance of multiple process cohorts. Rather than finding a theoretical truth, it is therefore focussed on achieving utility. The proposed tool has consequently been designed and implemented following a Design Science methodology [24], as described in Fig. 2.

The work was started by identifying an industry demand for a tool to enable process cohort comparison. This need is evidenced in a) the questions that our industry partners were asking and b) at least three recent case studies [2–4] reporting similar questions from industry. This industry need is therefore an input to the relevance cycle [25]. Furthermore, in Section 2, we discussed why current approaches are not well suited to address these questions. Based on this analysis and by building on knowledge from visualisation and process mining we identified ways to overcome the limitations of the existing approaches. These inputs come from the rigor cycle.

Consequently, the design cycle of this project was concerned with designing and implementing a tool for visual cohort comparison and a rigorous evaluation of this tool's utility for the described problem. In the evaluation of our tool we were guided by Munzner's nested model for visualisation design and evaluation [26]. We focussed our validation efforts primarily on layers two and three of the model, namely the data/operation abstraction design and the encoding/interaction technique design. Validating the domain problem characterisation is a long term endeavour and can only be done once a suitable problem solution has been developed. Algorithm design on the other hand was not really a concern as this work builds on existing algorithms rather than proposing new ones. The validation of the encoding/interaction technique design was complicated by the fact that existing approaches do not fully support the intended task of cohort comparison making a direct comparison unsuitable. We therefore opted for a descriptive validation by demonstration as described by Hevner [24]. In doing so we demonstrated that the proposed tool can be used to answer questions about two industry datasets with little effort. To validate the data/operation abstraction design and complete the relevance cycle we evaluated the resulting tool through a user study with two industry partners. The resulting feedback motivated the implementation of additional features, but also confirmed that the industry partners found the proposed tool useful for some analysis scenarios of interest.

In the rigor cycle, the abstraction of our solution, presented in the form of design principles, adds to the scientific knowledge on how to compare process cohorts. Since our contribution to knowledge draws on mature approaches from the field of visualisation and applies them to the novel problem of comparing process cohorts, it can be classified as an "Exaptation" as per Gregor and Hevner [27]. In the following sections we will describe which theories were used and how they were applied to address a set of new problems related to comparison of multiple process cohorts.

3.2. Detailed requirements

In order to design an approach and implement a tool, we proceed to discuss the requirements that stem from the given task. The aim of the work presented here is to facilitate the comparison of performance of multiple cohorts of cases in an event log. As was discussed in the related work, this aim already determines some high-level requirements.

Overall, our approach needs to cover both preparing the data for comparative analysis and presenting the data in a way that makes it easy for the user to interpret. Preparing the data requires support for two steps. Firstly, there needs to be a way to specify the cohorts the user wants to compare (R1). Secondly, computing metrics through which the performance of these cohorts can be compared is necessary (R2). Presenting the data in a way that makes it easy for the user to interpret can most likely be achieved through visualisation. As discussed in the related work, an integrated approach to present the data of more than two cohorts in one visualisation (R3) and to facilitate the comparison of this data (R4) is needed. Next, we break down these requirements in more detail both by making logical arguments and by using examples from the case studies [2–4] motivating this research.

Firstly, we identify ways to split an event log into process cohorts that can be compared. Generally, event logs contain information related to cases and activities. We chose to define cohorts on the case level, because a) many contextual factors affecting a case are constant over the time of its execution, and b) the occurrence and execution of activities that form a case is often based on these factors. For example, Suriadi et al. [2] analyse cohorts of insurance claims defined by the amount of payout and the total time it took to process these claims. Consequently, users should be able to define cohorts at the case level (R1.1). The approach should then be able to automatically split the event log so that cases are grouped for the subsequent computation of performance indicators and comparison (R1.2) to avoid lengthy data migration between different tools.

Secondly, we identify metrics to compare the specified process cohorts. The performance of cohorts can be compared by calculating different performance indicators using data in event logs both at the case and at the model level [28]. For this work we focus on comparisons at the model level using node and node-pair related performance indicators. Conceptually, these indicators describe attributes of activities (nodes in the model), such as the duration of an activity, and attributes related to pairs of activities that occur in sequence, such as the waiting time between two activities. We chose this focus as the model level is most likely to give interesting insights and can benefit from contextualisation of the performance data. For example,

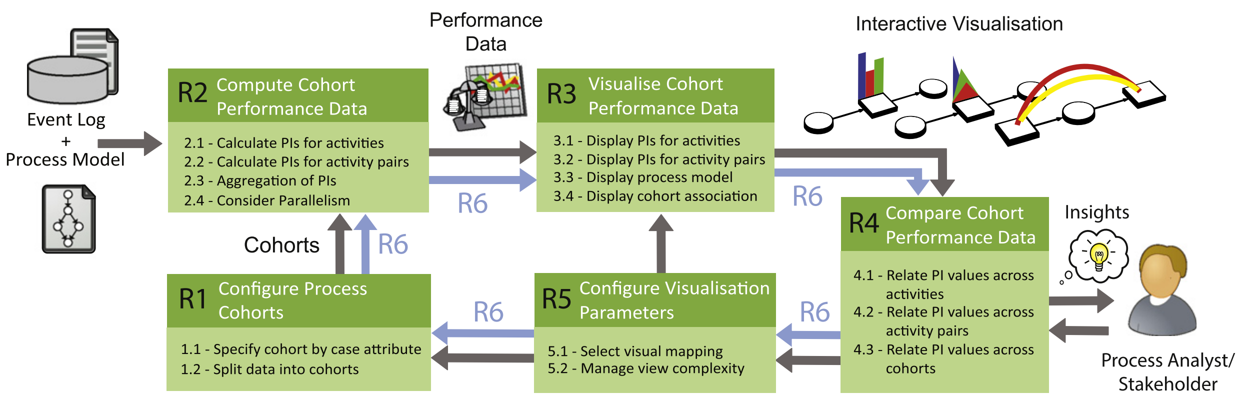


Fig. 3. Summary of the detailed requirements for an approach to visually compare process cohorts.

Partington et al. [3] use frequencies of occurrence between pairs of activities to compare processes cohorts. Therefore, the proposed approach needs to be able to compute these performance indicators on the level of activities (R2.1) and activity-pairs (R2.2). In addition, it should be possible to aggregate these values (R2.3) at higher levels of abstraction. This can help with the analysis of very complex processes, as for example was the case with hospital processes discussed by Partington et al. [3]. We furthermore identified the need to consider parallelism of activities in the performance analysis (R.2.4) in the related work.

Thirdly, we need to identify how these performance indicators can be presented to the user in a way that facilitates making sense of the data. Consequently, we need to find first of all a way to visually present the performance indicators for activities (R3.1) and activity-pairs (R3.2). However, the performance indicator values have limited meaning if presented on their own. The contextual variables that give meaning to them are the cohort, and the activity or activity-pair that these values describe. This means that these three values need to be shown in combination to enable users to understand them. Furthermore, in order to analyse why differences exist between process cohorts, the user also needs to see these differences in the context of the interdependencies between activities. For example, a bottleneck in a process can be found by finding an activity that takes much longer than activities preceding and following this activity. The dependencies and ordering of activities can be represented using a process model [1]. This contextual information will help the user to better understand problems in the process. Consequently, it is important to contextualise the performance data by visualising it together with the process model (R3.3) and clearly showing which cohort it describes (R3.4).

In relation to the specific task at hand, i.e. comparing process cohorts, it should be possible to display multiple sets of data values at once. Depending on the specific scenario, the analysis might then require the user to compare performance indicators across activities or activity pairs to identify trends and bottlenecks (R4.1, R4.2) as discussed above, compare values across cohorts (R4.3) or even both. For example, Partington et al. [3], compare “hours till inpatient care” and “hours in inpatient care” by aggregating time from duration and waiting times along parts of the process model and then compare these values between multiple cohorts.

An additional set of requirements (R5) is rooted in the generic tasks users perform to explore and understand visual data [29]. A number of visual mapping parameters (such as data normalisation modes) should be accessible to the user to suit their scenario and individual preferences (R5.1). Furthermore, displaying a large dataset in an integrated way will lead to high visual complexity, therefore the user should be able to reduce the visual complexity (R5.2).

A final requirement (R6) stems from the way these tasks are integrated. Not only does such an integration remove tool chain interoperability issues as discussed by Partington et al. [3], it is also the only way to enable truly interactive exploration of process data. Interactivity is important for knowledge generation in visualisation [30] and its effects on knowledge generation processes and outcomes have been demonstrated even for small interaction latencies [31]. The need to repeat steps of data filtering and to iteratively adjust the specification of cohorts in order to find meaningful cohorts to compare has also been reported [3]. Fig. 3 summarises the requirements that have been derived from the task of visually comparing process cohorts as discussed above.

Building a tool that meets these requirements should enable users to intuitively compare process cohorts based on event data. The next sections discuss design decisions made to satisfy these requirements in turn.

3.3. Computation of performance-related statistics

To compute the performance statistics, we generally follow the algorithms presented by Adriansyah [28]. We compute an alignment between a given event log and a Petri net process model [32] using an existing ProM [33] plugin (“ Replay a Log on Petri net for Performance/Conformance Analysis” [6]). Given this alignment we can reconstruct the movement of tokens through the Petri net and create a graph of dependencies between events in the event log, as

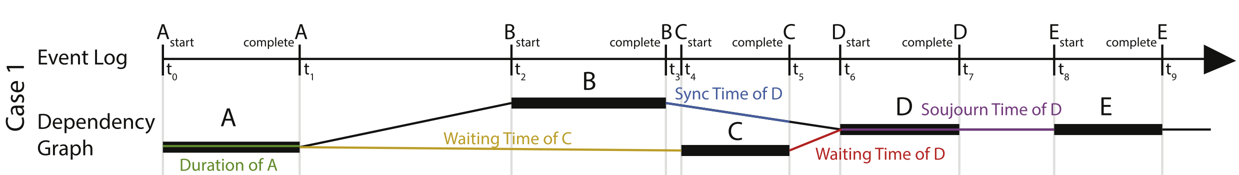


Fig. 4. Computation of performance indicators for case 1 in an event log. Each edge in the graph represents a time interval. Some performance indicators can be the sum of several edges.

shown in Fig. 4. Conceptually, each edge in this graph then represents a time interval. Therefore these edges can be used to calculate timing statistics such as durations and waiting times of activities using only addition and subtraction. This satisfies requirements R2.1 and R2.2. In addition, since process structure is encoded in the Petri net, this approach also deals with concurrency of events in the log. For example, without the graph, the waiting time of C would be incorrectly calculated as $C_{start} - B_{complete}$. The approach therefore also satisfies requirement R2.4.

Our approach differs from that of Adriansyah [28] in two ways. Firstly, the order of individual steps in the algorithm is reversed to reduce computational overhead and enable interactive data exploration. Adriansyah's approach assumes a fully prepared event log for analysis. That means that to analyse two cohorts, the original log would be split into two event logs and then each log is aligned with the model before computing performance statistics. In order to compare two different cohorts, the original log would have to be split again, and both alignments with the model and the performance indicators would have to be recomputed. Our approach first aligns log and model once, then computes performance indicators once. Then, instead of statistically summarising the calculated performance indicators at the time they are computed, we store each individual value (i.e. edge in the dependency graph) for each case contained in the log, using an online analytical processing (OLAP) approach. This approach enables users to repeatedly split performance data into cohorts and quickly aggregated summary statistics for each cohort from the individual values. Consequently, while the alignment may still take a long time to compute (as in the approach by Adriansyah), we only have to compute it once and can then calculate the performance data for each new cohort in real-time and with much less overhead. This enables the dynamic aggregation of performance statistics for user specified cohorts (at runtime), satisfying R1.2. Secondly, for hierarchical Petri nets we aggregate values at higher levels of the process by adding up the time intervals of connected edges that together represent a sub-process. Doing so enables the display of performance indicators at multiple levels of abstraction of the process, as required by R2.3.

An underlying assumption of this approach is that the structure of the process model used for the computation does not change, as the organisational process represented by the model should be static over the course of analysis. However, analysts could be interested in adding levels of hierarchy to aggregate performance data for different segments of the model. In principle, this is supported by the approach, as the underlying process structure does not change and only the aggregation step would have to be repeated for new subprocesses, which generally takes little time.

3.4. Visualisation design

Our requirements discuss that all cohort performance values including their association to a cohort, an activity in the process and any dependencies with other activities should be presented in one view. We therefore need to encode performance statistics and contextual variables pertaining to these statistics, such as process cohorts and activity they belong to, in different visual dimensions of that view. In general, data can be encoded in one of seven visual variables, such as shape, colour, size, orientation, brightness, texture and position of a visual element [34]. The performance data exists in the form of counts, such as frequency of occurrence, and time spans, such as activity duration and waiting time. In both forms it is continuous and can potentially have a large range of values. The largest capacity of values that can be visually distinguished is in the size variable [35], making size an ideal dimension to encode performance data in.

Existing approaches to performance visualisation therefore often encode performance data in the visual parameters of process model elements and consequently present a process model, such as the size of an activity. However, such an approach can only encode one dimension of data. For the comparison of two cohorts one can instead encode the difference between these cohorts to work around this limitation, but we want to be able to compare more than two cohorts. Bar charts are a common approach to visually compare multiple values encoded in size along the same dimension. However, bar charts do not provide a way of encoding dependencies the way a process model does. We therefore combine principles from both visualisations, by *overloading* the space of the process model with additional visual elements [36], one for each cohort per activity. However, we encode the data values in the size of the visual elements in a dimension that is orthogonal to the two dimensions used for the process model. Using this third, orthogonal, dimension of height exclusively to visualise the performance statistics values means that all values can be seen relative to the ground plane. Our encoding uses the principle of superimposition as all data is encoded along one dimension (i.e. "height") and therefore facilitates comparing values in the same way a bar chart does [36], supporting both global comparison of values and comparison of local values within a model wide overview context [29]. At the same time this approach minimises occlusion and layout issues and maximises perceptual popout, enabling the user to view the data while not losing any of the data's context [18].

As mentioned in the requirements earlier, the data also requires context information to understand it properly. The visual elements are drawn connected to the model elements they describe, using the principle of "connectedness" [35] to signify which piece of data belongs to which activity or pair of activities in the process model. In

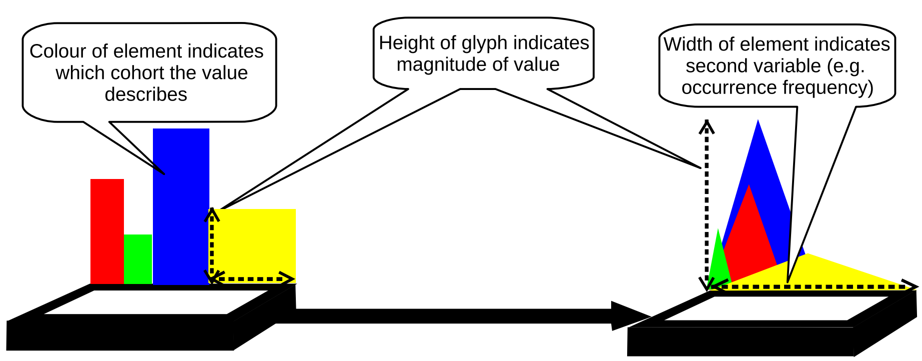


Fig. 5. Visualisation of performance statistic values relative to red cohort value using colour coded bar charts (left) and triangle charts (right) on top of activities.

addition, there are always multiple but few cohorts (2–8), which can be considered a nominal variable in terms of data. The human visual system can distinguish colours most effectively, especially a range of 7–10 different colours [35]. We therefore chose colour to represent visual glyphs of one cohort satisfying requirement R3.4.

In the case of performance statistics belonging to an activity we draw visual glyphs, previously introduced by Pini et al. [5], on top of the activity. The two main techniques that are used to compare relative values between cohorts for each activity are bar charts and triangle charts (see Fig. 5). The first technique shows the value of the first variable for a transition for each cohort as the height of a coloured bar and the relative value of a second variable as width of each bar segment. This enables displaying performance indicators such as the duration of an activity together with contextual information, such as how often the activity occurred. The second technique can show two variables as width and height of an isosceles triangle to the same effect.

For activity pairs, we utilise an arc visualisation, inspired by an iso-line approach taken from sun radiation visualisations [37]. Arc visualisations have traditionally sought to show connectedness relationships between diagram components [38] and in our approach, the arc starts at one activity of the respective activity pair and ends at the other. The height of the arc signifies the magnitude of the visualised value as shown in Fig. 6.

To prevent cognitive overload, users are provided with ways to filter out and aggregate information as forms of complexity management. Firstly, due to the use of a three-dimensional space for visualisation the user can move and rotate the camera in all three dimensions. This way the user can focus the view on regions of interest in the process model so that uninteresting data is outside of the view region or to avoid occlusion issues. Secondly, users can see aggregated data using hierarchy in the process model and filter out irrelevant data by selecting individual model elements or cohorts of interest. These features satisfy requirements R5.1 and R5.2.

Overall these principles have been chosen to enable an intuitive visualisation and comparison of performance data from multiple cohorts.

3.5. Design principles

We have shown in the related work, that support for process cohort comparison is still fragmented and no existing approach satisfies all the requirements we have identified in the previous section. We see this as an indication that general guidance on designing solutions that address this problem may be needed. While we have discussed specific design decisions that were made in order to develop a system that addresses all requirements, we also want to abstract out general design principles from this solution that will provide guidance in the development of other tools for process cohort comparison. These principles explicate abstract features of form and function inherent in the design of our instantiation and relate these features to the identified requirements in order to provide guidance for the development of further artifacts of the same class. They thereby constitute explanatory design theory as per Baskerville and Pries-Heje [39].

Fundamentally, comparing the performance of process cohorts requires preparing and presenting the performance data. Preparation usually includes splitting the event log into cohorts, aggregating event data to a sensible level of detail and computing performance indicators for each cohort. As reported by Partington et al. [3] this process is often repeated as cohorts have to be adjusted and the focus of interest shifts between different parts of the process. We discussed earlier that for gaining insights into data interactive data exploration is important [30]. In order to make data exploration interactive, we needed to reduce a) the time users spend moving data back and forth between different tools and b) the time spent recomputing alignments between log and model. To this end, we needed to integrate both data preparation and data presentation. We therefore posit that a tight integration of data preparation and presentation should allow users to interactively explore their process data and in turn compare process cohorts more effectively:

- **DP1**: Integrate data preparation and data presentation to enable interactive data exploration.

To address requirements relating to the computation of process performance metrics for cohort comparison in more detailed we propose three additional design principles:

- **DP2**: Enable users to specify concurrent activities to correctly compute performance indicators in the presence of parallelism
- **DP3**: Enable users to define cohorts interactively to test hypotheses
- **DP4**: Present performance data at multiple levels of abstraction to allow users to gain an overview of the data and drill-down for details-on-demand.

Firstly, process data in the form of event logs usually serialises the entire process and loses dependency and concurrency information in doing so. However, this information can often be relevant in performance analysis, for example when computing waiting times between different parts of the process. An approach for cohort comparison should therefore enable the user to provide domain knowledge, for example in the form of a process model, so that the analysis is aware of the underlying process structure. Secondly, as has been

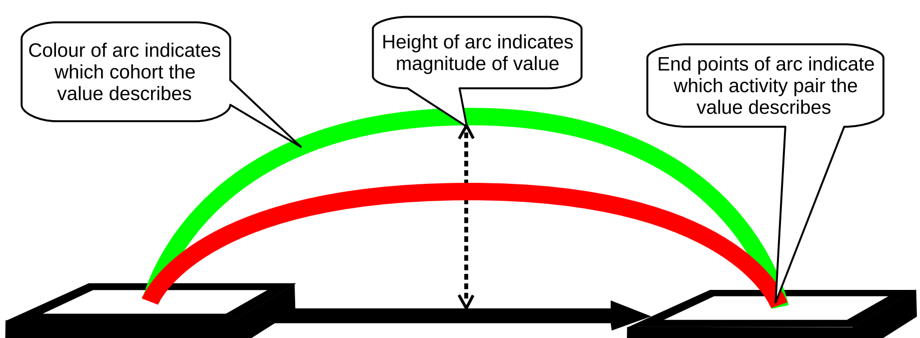


Fig. 6. Visualisation of performance statistic values using colour coded arcs between activity pairs.

Table 2
Summary of design decisions that implement design principles and satisfy our requirements.

DP1	OLAP approach enables interactive data preparation and presentation	R6
DP2	Align event log with process model for computation of node and node pair related PIs in the presence of parallelism	R2.1
		R2.2
		R2.3
DP3	OLAP approach enables splitting data into cohorts, activities and PI types	R1.1
		R1.2
DP4	OLAP approach can aggregate data on multiple levels of abstraction interactively	R2.3
DP6	Visual glyphs from Pini et al. [5] enable presentation of node related PIs	R3.1
	Use of arcs enables presentation of node pair related PIs	R3.2
	Show flat process model	R3.3
	Use of colour to show association of visual glyphs with cohorts	R3.4
DP5	Node related PI visual glyphs and node pair related PI arcs are connected to nodes in process model, superimposed in shared vertical axis for easy comparison	R4.1
		R4.2
	Values for different cohorts share vertical axis, related values are located nearby	R4.3
DP7	Different glyphs and normalisation modes can be used to customize visualisation	R5.1
	Navigation of 3D space, zoom and filter can be used to manage view complexity	R5.2

discussed above, the ability to interactively define cohorts and split the event log is important for cohort comparison. Finally, real-world processes can be very complex, therefore comparing them at a high level and then drilling down into the details that are of interest can help to make the comparison more manageable. Partington et al. [3], for example, discuss how they aggregated metrics for several higher-level stages of the hospital process to deal with the complexity of their process model for analysis. We therefore argue that it is necessary that an effective approach to compare process cohorts needs to enable users to compare their cohorts at different levels of detail.

When comparing process cohorts using performance metrics the presentation of cohort performance should make the identification and understanding of differences between multiple cohorts easy. We present three more design principles that can be used to guide the design of process cohort comparison approaches to do so.

- **DP5**: Visualise cohort performance data in one integrated view to facilitate comparison
- **DP6**: Present related data in orthogonal dimensions to enable contextualisation of the data
- **DP7**: Enable users to navigate and filter the visualisation to manage view complexity.

Firstly, we propose that all the relevant data attributes for cohort performance comparison need to be presented in one integrated view. We reiterate that multiple data attributes (such as association of cohort, activity, performance indicator) are integral to the analysis task and juxtaposing multiple complex visualisations scales badly and makes the analysis harder for the user [17]. A solution to this problem is therefore an integrated visualisation that shows all the relevant data in one view. Secondly, presenting multiple data attributes in one view is not trivial as this requires that multiple sets of values (i.e. the average waiting time in each cohort for one specific activity) a) can be viewed at the same time, b) can be clearly distinguished from each other and c) can be distinguished from unrelated values (such as the waiting times for other activities) so that no visual overload occurs. One way to ensure that different values can be clearly separated and related values can be clearly compared is to use a separate visual dimension for each attribute. This way *superimposition* can be used to ensure good comparability of values and preattentive processing can be used to filter out unrelated data. Lastly, in addition to presenting the data in a cognitively efficient way, the user should also be provided with ways to manually manage view complexity.

Together, these design principles inform the design of an approach to compare process cohort performance effectively. Table 2 shows how the requirements identified earlier are met by design decisions that were made and which design principles these decisions implement. Next we will demonstrate that the designed approach for cohort comparison is feasible and has been implemented as a research prototype tool by discussing our implementation.

4. Implementation

As a proof of concept the computation of performance indicators as described above, as well as the visualisation of the performance indicators, have been implemented in two separate plug-ins for ProM [33]. The “Process Profiler 3D: Generate Performance Profile” plug-in calculates performance statistics for a given Petri net by

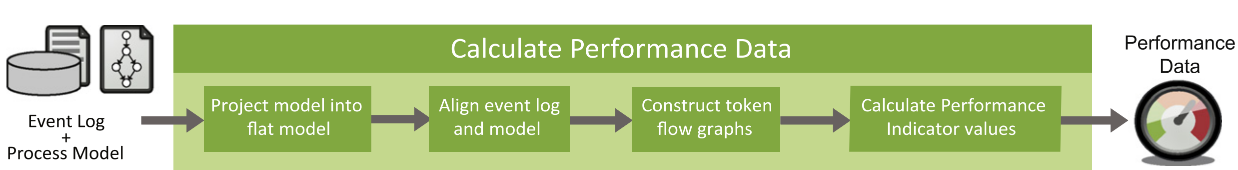


Fig. 7. Process Profiler 3D: Generate Performance Profile - ProM plugin overview.

RowID	CaseID	CohortID	NodeID	LabelID	Value
1	Case 1	Cohort 1	Activity A	frequency (count)	1
2	Case 1	Cohort 1	Activity A to Activity D	frequency (count)	1
3	Case 1	Cohort 1	Activity A to Activity D	waiting time (timespan)	1 day 4 hrs
4	Case 2	Cohort 1	Activity A to Activity D	waiting time (timespan)	18 hrs 31 mins
5	Case 11	Cohort 1	Activity A to Activity D	waiting time (timespan)	19 hrs 2 mins
…	…	…	…	…	…
166	Case 37	Cohort 2	Activity A to Activity D	soujourn time (timespan)	6 hrs 5 mins
167	Case 37	Cohort 2	Subprocess B	duration (timespan)	5 hrs 40 mins

Fig. 8. An example of the performance data that is stored in the data cube. To compute the summary statistics we group rows by A) cohort B) node or node pair ID C) PI label and then D) summarise over the grouped values.

replaying a given event log (Section 4.1). The "Process Profiler 3D: Visualise Performance Profile" plug-in visualises the data in various ways that support the identification of performance differences between process cohorts (Section 4.2).

4.1. Creation of a performance statistics enriched net

For the computation of performance data, a number of steps are performed which are discussed in this section. Fig. 7 provides a high level overview of the operations the plugin performs.

For this work we assume the presence of an event log that contains at least timestamps for the completion of all activities. Furthermore, we assume the existence of a process model which has a high fitness for the event log in the form of a Petri net. A misaligned process model can result in the computation of incorrect performance metrics. While this assumption is not trivial, several techniques are available to manually create or automatically mine such models from event logs [1]. Similarly, hierarchy in process models can be achieved either by automatic decomposition or manual editing in process modelling tools such as WoPeD [40]. For the replay of logs on hierarchical Petri nets the hierarchical net is projected into a flat net. This projection is performed automatically by recursively replacing each activity with its corresponding subnet. Then a replay of the event log on the process model is performed using the replay plugin of ProM to create an alignment between the Petri net model and the log. Using the resulting alignment, the flow of tokens through the Petri net is reconstructed. The reconstruction enables the identification of dependencies between the events in the log and is used to calculate the required performance indicators. We calculate the performance indicators by recording a token path for each move of a token on the Petri net from one place to another. Each token path instance is linked to preceding token path instances that created the tokens it consumed and stores the event in the event log that it corresponds to. Each token path instance also has timestamps attached to its beginning and end based on when the token in the start place was created and when it was created in the end place. Each activity in the process is now represented by one or more consecutive token path instances that each represent a time span. Consequently, each case of the process is represented by a directed graph annotated with timestamps. We can therefore compute timespans between different activities by traversing this graph, which results in a composite path with a start and complete timestamp. The time span that is represented by this composite path is one performance indicator value. The resulting performance indicator values are then stored in the data cube.

The interactive comparison of process cohorts is supported by the use of data warehousing techniques [41], in the form of a relational online analytical processing (ROLAP) datacube, to store individual performance indicator values and attributes that describe their context in a database. Fig. 8 shows an example of the data that is stored in this datacube. As shown, there exist zero or more performance indicator values for each case in the database. An example of a database row for a performance indicator value would be a numeric value representing the waiting time for one activity executed in one case of the event log, the case identifier of the case, the identifier of the activity and the label identifying the semantics of the stored value, i.e. identifying the value as a waiting time. The identifier of the cohort that case is currently assigned to is also stored and updated whenever the cohort specification changes. This approach enables us to dynamically split the dataset into cohorts based on these attributes, which is known as "slice&dice" (satisfying R1.2) and to aggregate statistical summaries of the indicators (such as minimum, maximum and mean) at runtime, which is commonly referred to as "roll-up" operation (satisfying R2.3). Furthermore, performance values for activities are computed and stored at each level of the hierarchy provided by the process model, so that the user can view performance summaries at different levels-of-detail. By using these operations, users can interactively explore the dataset

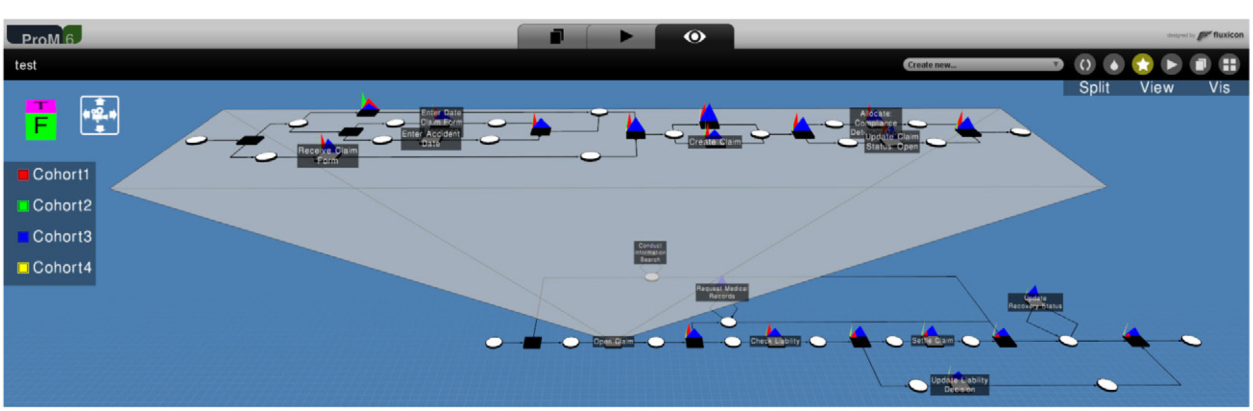


Fig. 9. Process Profiler 3D: Visualiser - ProM plugin.

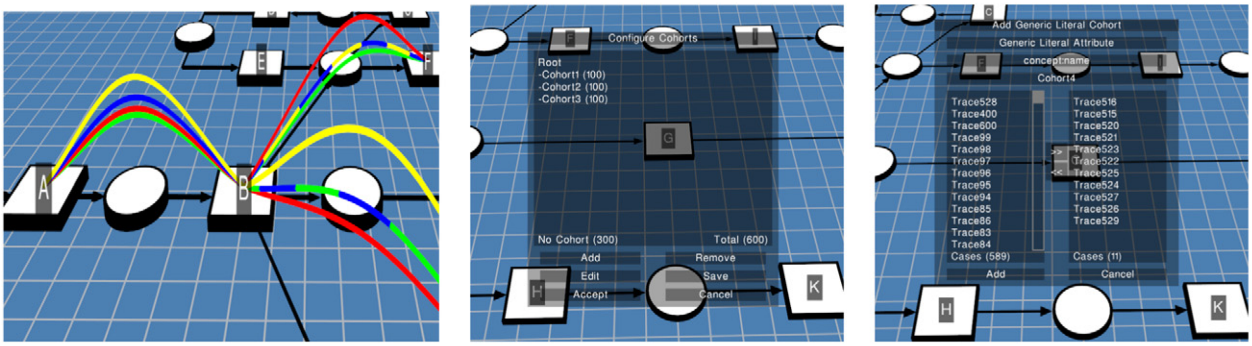


Fig. 10. Left: visualisation of differences in waiting time between transitions; center and right: specification of process cohorts based on case level attributes of the event log.

and test their hypotheses rather than having to recompute the performance statistics for each different question they would like to investigate. In addition, since the performance data is computed and stored independently of the layout of the process model, users can modify the model layout and still visualise the data without having to recompute it. While it would be possible in principle to let users insert additional levels of hierarchy into the model interactively by re-aggregating performance data, our implementation currently does not support this, as adding model editing features would have significantly increased the complexity of the implementation. The complete implementation of the algorithm can be found in the code repository at https://svn.win.tue.nl/repos/prom/Packages/ProcessProfiler3D/.

4.2. Performance data visualisation

The performance data is visualised in widgets on top of the (hierarchical) Petri net. Fig. 9 shows an example of cohort performance data visualised using the performance profile visualisation plugin for the ProM framework. In order to reduce layout and overlap issues when visualising this data, a three-dimensional approach was chosen in which the Petri net is shown in two dimensions and performance data is visualised in the third dimension. This approach enables the user to handle overlap or information overload by changing the perspective from which the data is shown. As suggested by the design principles (see Section 3.5), the plug-in provides functionality for visualising cohort performance data, managing view complexity and defining process cohorts.

Currently several visualisation techniques are available to compare the statistics data calculated for each transition including two-dimensional barcharts and triangle glyphs. Furthermore, statistical data that relates to connections between transitions, such as waiting times between activities, can be visualised using the arc visualisation technique described in Section 3.4 (see Fig. 10). If the values of some cohorts are close and overlap, we merge the arcs to avoid visibility issues. In that case the arc will be drawn with coloured stripes for each cohort that it is representing.

In order to manage the complexity of the visualisation, users of the plug-in are provided with options to abstract, filter and normalise data. Firstly, abstraction is provided on the level of the Petri net. If the net is hierarchical the user can view the aggregated performance statistics of a transition and open up the subprocess to view data at a lower level (see Fig. 9). Secondly, the user can choose to hide visualisations concerning specific transitions, arcs and cohorts. Thirdly, three normalisation modes are provided to normalise visualisation elements at the level of a transition, sub-process or the entire model. Additionally, visualisation arcs can be normalised by cohort, as discussed in Section 3.4.

Process cohorts can be specified by the user at runtime. To specify a cohort, the user selects a case attribute available in the event log and then specifies which cases are included in the cohort by selecting the set of attribute values that should be included (see Fig. 10). The plug-in then automatically aggregates and displays the performance statistics for the specified process cohorts.

Overall, the two ProM plug-ins together implement the framework for visual process performance comparison that has been described in Section 3. They enable the comparison of event log cohort performance in a quick and intuitive visual manner.

5. Validation

In this section, two evaluations of the utility of the prototype tool will be presented. We define utility as *the tool's ability to enable users to compare the performance of multiple process cohorts interactively*.

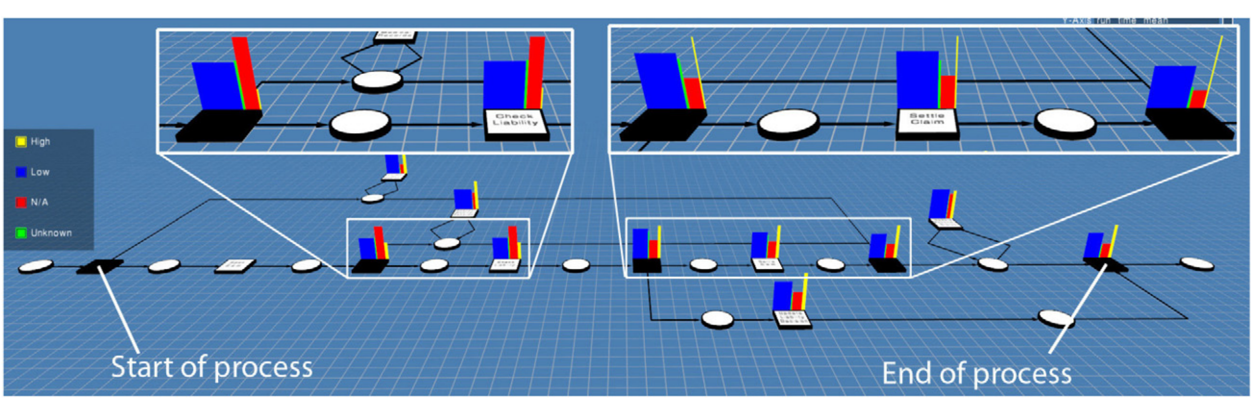


Fig. 11. Analysis of case runtime by injury severity grouped into four cohorts in ProcessProfiler3D. A trend change can be observed after the Check Liability activity in the process for both the injury severity "Not reported" cases (red) and the "High" injury severity cases (yellow).

Firstly, this utility will be demonstrated by showing how the tool can be used to analyse two industry datasets. Secondly, an in-depth user study that was performed with the practitioners from two organisations will be discussed. These evaluations were performed in the context of a larger research initiative on identifying impediments to insurance claims processing which is supported by Queensland Motor Accidents Insurance Commission and the Queensland Government. The scope of the analysis reported in this paper is limited to the evaluation of the utility of the prototype and the comments provided by the participants in the user study are also limited to the impressions that the participants have about the prototype tool.

5.1. Demonstration with industry datasets

As a first step in the validation of the prototype tool, two industry data-sets were analysed using the different functionalities supported by the tool. The first event log was provided by the Nominal Defendant (ND), an organisation that provides insurance compensation for injuries resulting from negligent driving of unidentified or unregistered motor vehicles in Queensland, Australia. The event log contains data relating to 964 insurance claim cases handled by the Nominal Defendant between 2006 and 2015. The log consists of 25,571 events with 66 distinct activities. The median claim took 23.3 months to handle from start to finish. Another event log was provided by the Royal Automobile Club of Queensland (RACQ), a licensed Queensland CTP insurer. The process for claims handling is similar to the Nominal Defendant on a high level, but the details of claims handling differ. The event log contains data relating to 1091 claim cases handled by RACQ between January 2012 and July 2015 and consists of 44,786 events with 65 distinct activities. The median claim took 16.7 months to handle from start to finish. Both logs also include several case attributes that describe the context of the handled claim, such as the severity of claimant injury, plaintiff and defendant law firms involved in the case and anonymised information about the claimant. After an initial cleaning of the log data, hierarchical Petri net models of both claim handling processes were constructed from these logs. We validated the models with the Nominal Defendant and RACQ by explaining the model to their process managers and confirming that they represent the processes correctly. We analysed both logs to answer multiple questions for both industry partners. To demonstrate the utility of the tool, two examples taken from this analysis will be discussed. A question of particular interest to the Nominal Defendant was how the severity of the injury affects process performance. RACQ wanted to know whether there are differences in process performance based on the region the solicitor's offices are located in.

The first question was investigated by splitting the dataset into process cohorts based on the reported injury severity of the claimant. This resulted in four cohorts: "Not reported" (235 cases), "Low" (666 cases), "High" (25 cases) and "Uncertain" (38 cases). The "Not reported" cohort contains cases for which the dataset did not contain injury severity scores. Cases with "Low" injury severity involve injuries that are generally considered not life-threatening, ranging from minor lacerations to multiple bone fractures. Cases with "High" injury severity involve injuries that are considered life-threatening or fatal. The "Uncertain" cohort contains cases where it has not yet been possible to judge the impact of an injury, e.g. in the case of a severe head injury. The average case runtime per cohort was then visualised on top of the high-level process model, normalised per activity (see Fig. 11). It can be clearly seen that cases with high injury severity overall take longer than cases with low injury severity, as indicated by the bar chart on the last activity. Interestingly, we can also clearly observe that cases with the injury severity being either "Not reported" or "Uncertain" take more time than any other cohort in the early stages of the process and then proceed much faster towards the end of the process.

The second question was investigated by splitting the event log into four cohorts, based on the postcode of the claimant solicitor's offices. To this end, the postcodes were grouped into four geographic regions that are meaningful to the industry partner. The first cohort includes Brisbane's urban and suburban area (394 cases). The second cohort includes the remainder of South East Queensland including the Gold Coast and Sunshine Coast regions (186 cases). The third cohort contains the remainder of Queensland (67 cases) and the final region groups all solicitors with offices outside of Queensland (22 cases). When the average waiting time between all activities is visualised per cohort (see Fig. 12), it can be seen that no clear differences emerge and only two arcs stick out of the rest of the visualisation. On closer investigation, these two arcs represent only three cases each, so they constitute outliers, rather than significant trends. From this, we can conclude that there are no significant differences visible for cohorts based on the location of claimant solicitor's offices.

Overall, we successfully used the tool to investigate multiple questions that were of interest to our industry partners and discussed two of them as examples. The proposed tool exposes patterns in the data that would have remained unnoticed using traditional spreadsheet-based approaches. Moreover, the results can be related much better to the actual process as trends over time (in this case not relating to absolute time spans, but rather progression through the process) can also be discerned. These examples involving real-world data-sets demonstrate that the tool can indeed be useful for an objective comparison of process cohorts.

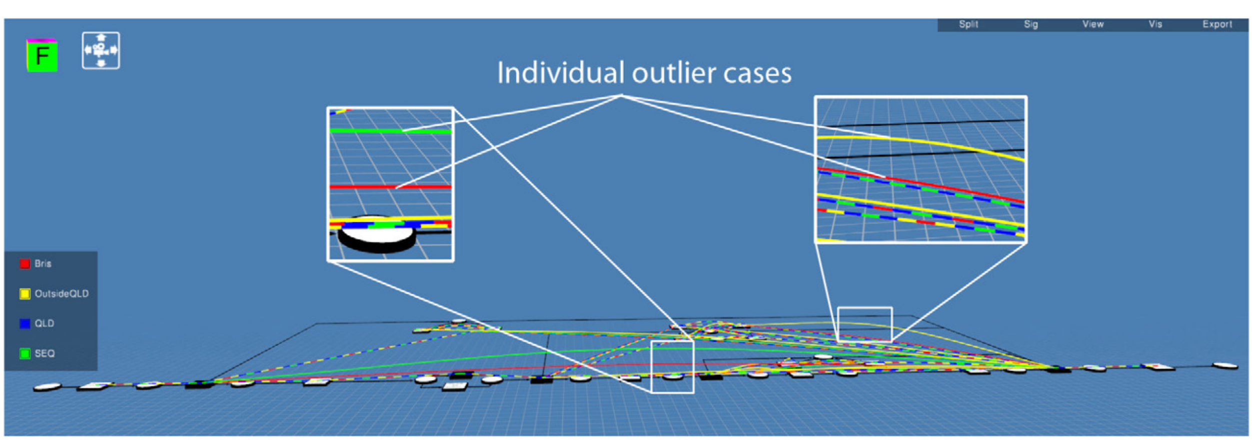


Fig. 12. Analysis of the average waiting time by claimant solicitor's office location grouped into four cohorts in ProcessProfiler3D. Most arcs are merged for all cohorts, indicating that there is no significant difference in waiting times between the cohorts. A few individual case outliers can be seen sticking out.

5.2. User study

While the demonstration provided some evidence that the proposed approach to comparative analysis works, demonstrating the relevance of the tool in a real-world setting is an important part of design science. To gather in-depth feedback on how the tool would perform in such an environment, a study was performed with our two industry partners (ND and RACQ, see Section 5.1).

In this study we investigate ease-of-use and usefulness of the proposed tool for industry practitioners, as they are major factors influencing adoption of technology [42]. We adopt the definitions of Davis [42], who defines ease-of-use as "the degree to which a person believes that using a particular system would be free of effort" and usefulness as "the degree to which a person believes that using a particular system would enhance his or her job performance". We therefore wanted to identify individuals in the organisations that would use such a tool in their work. We used a purposeful sampling strategy, snowball sampling [43], to identify 12 individuals involved in process management and execution in both organisations across a variety of roles, including team leaders, managers, business intelligence and data analysis roles. The tool prototype was demonstrated to these individuals by visualising log data from each organisation and letting the stakeholders propose configurations of visualisation parameters to gain insights from the data. The purpose of this demo was to give them an understanding of what kind of insights can be gained with the proposed tool and how much effort would be involved in using it. Afterwards their opinions and feedback on the tool were collected with a set of Likert-scale questions modified from the perceived ease-of-use and perceived usefulness scales of the TAM questionnaire [42] and a semi-structured interview with open-ended questions. The questionnaire covers two topic areas: the perceived ease-of-use of the visualisation (PT1–PT10) and the perceived usefulness of the prototype tool (PT11–PT16). Answers on a 7-point Likert-scale ranged from "Strongly Disagree" (1) to "Strongly Agree" (7).

Both measurement scales showed high reliability, with Cronbach's alpha coefficients of 0.832, 0.934 respectively. The perceived ease-of-use was overall rated positive, whereas the perceived usefulness was rated less positive. This observation has been confirmed by testing whether responses are significantly different from the neutral response using a one-sample Wilcoxon Signed Rank test (see Table 3). A precondition for this test is that each pair is chosen randomly and independently. While a purposeful sampling approach was used to identify the participants, the preconditions for the test have not been violated, as each individual was engaged in a separate time slot and no screening of participants was performed by the research team. There is no indication that the sample of participants engaged in this way would differ significantly from the target population (i.e. all process data analysts in any organisation). The results show that for 13 of the 16 questions the responses are significantly different and on the positive side. The medians for PT1-PT10, however, are higher than those for PT11–PT16. While these results should not be over-interpreted due to the low number of participants, we were able to confirm these sentiments in the qualitative analysis of our interviews with the same participants as well.

The interviews were transcribed and the comments of the 12 participants were coded independently by two members of the research team to categorise the comments. A meeting to converge on shared categories found that, despite slight differences, the categories of both coders mapped well to each other. As a result three main categories of statements made by the participants were identified: positive comments about the tool, negative comments about the tool and suggestions for improving the tool. The categories of the second coder also led to the decision to split both the positive and negative comment categories into ease-of-use and usefulness related subcategories.

Regarding the ease-of-use of the tool, the majority of participants stated that it helped them to *quickly identify anomalies and trends in the data*. Some participants felt that the visualisation *integrates well with their mental model of the process* and therefore enables an easy comparison with existing process knowledge. The participants generally found the visualisation *impressive and appealing* and many participants stated that *overview, drill-down and the normalisation modes were helpful* in gaining insights into the data. While some participants felt that the tool would be *easy to learn* and *easier to use than spreadsheets*, a few participants thought the 3D aspect could make the tool difficult to use and would require training. There were also concerns that business users might be unfamiliar with the terminology used for visualisation parameters and the interface being more complex than some existing (but less powerful) tools. Two specific issues with the visualisation were mentioned. Firstly, two participants noted that the arc visualisation can be hard to read when it gets crowded. Secondly, a participant pointed-out that bolder colours used to indicate specific cohorts might affect the perception of their dominance over the process model. A summary of the themes concerning the ease-of-use of the tool is provided in Table 4.

Talking about usefulness, the majority of the participants thought the tool can be used to *draw relevant insights from process data* if data and process models of high quality are available. One person pointed out that the tool *could improve process analysis by reducing the role of instinct and guesswork in analysing the data*. Many participants agreed that the tool *enabled them to quickly identify areas that required further analysis* and helped them to confirm or question pre-conceptions

Table 3
One-sample Wilcoxon Signed Rank Test to determine significant difference from neutral response. Answers differ statistically significantly from the neutral response at $p < \mathbf{0.05}$.

PT	Question	Z	Med.	Sig.
1	I found the visualisation easy to learn.	3.140	6	**0.002**
2	I found exploring the data using the visualisation was easy.	2.818	5	**0.005**
3	I found the parameters of the visualisation system easy to understand.	2.790	6	**0.005**
4	I found it easy to combine parameters to create visualisations.	2.842	6	**0.004**
5	I found it easy to see the relationships between cohorts in the visualisation.	2.980	5.5	**0.003**
6	I found the data in 3D visualisation to be hard to track when moved.	−0.426	3.5	0.670
7	I found the organisation of the data in the visualisation to be clear.	3.035	6	**0.002**
8	I found it straightforward to understand the visualisation tasks presented to me by the research staff.	3.213	6	**0.001**
9	I found it easy to compare a cohort with other cohorts.	3.134	6	**0.002**
10	I found it easy to view a subset of cohorts from the set of cohorts presented.	3.002	6	**0.003**
11	Using the 3D visualisation would enable me to accomplish tasks more quickly.	2.326	5	**0.020**
12	Using the 3D visualisation would improve my job performance.	2.333	4.5	**0.02**
13	Using the 3D visualisation in my job would increase my productivity.	1.667	4	0.096
14	Using the 3D visualisation would enhance my effectiveness on the job.	2.157	5	**0.031**
15	Using the 3D visualisation would make it easier to do my job.	1.035	4	0.301
16	I would find the 3D visualisation useful in my job.	2.156	4.5	**0.031**

Bold entries highlight gaps in current tool support for comparative process cohort performance analysis.

Table 4
Themes surrounding the ease-of-use of the tool.

Positive	Negative
• Visualisation is impressive and appealing • Would be easier to use than spreadsheets and easy to learn • Helped quickly and easily identify anomalies, outliers, differences and trends in data • Overview, drill-down and normalisation modes are good to gain insight into data • Visualisation integrates well with mental model of process	• User interface not (business) user-friendly • Tool has more overhead than simpler tools • 3D is perceived as difficult and would require training • Arc visualisation sometimes hard to read • Colour codes for cohort can affect perceived dominance of a cohort

about process performance for different cohorts. The various features of the tool, such as providing a quick overview of all data, drill-down into the details and normalisation modes, were perceived as *good for analysing and managing complex processes*. Some participants also came up with additional use-cases for the tool that we had not considered. Firstly, it was suggested that the interactive view could *improve process understanding of stakeholders* as it was easier to understand than paper printouts of the process. Secondly, some participants suggested that the visualisation could be used to justify process changes by *demonstrating differences in process performance* to other stakeholders.

Despite these overall positive comments, three participants were uncertain about the business value of the application, due to perceived limitations of the tool. One problem mentioned is that the data that is currently recorded and available for analysis often does not provide enough detail to create interesting and business relevant insights. Consequently, while the participants saw the usefulness of the visualisation aspect in the demonstration of the tool, some of them did not find the insights gained from the presented analysis scenarios particularly surprising from a business perspective. Another issue is the *availability of a good process model* that fits the event log while being human-readable. Some participants also mentioned that managing *simple processes may not require the level of sophistication provided by the tool* and is already covered by simpler tools.

The specific skills required to use the tool may also create a barrier to using the tool for business users. For example, four participants were concerned that people without sufficient process modelling expertise or domain knowledge would not be able to understand and use the tool. An interesting discussion, as to whether the tool's analytic capabilities would be more beneficial for managers, process participants or both, emerged from these points. However, many of the issues constraining the usefulness of the tool are contextual rather than inherent to the tool itself. Accordingly, given high quality data on a complex process and for a business user with process knowledge, the tool would likely be very useful. These issues are all summarised in Table 5 below:

The participants also suggested a number of features that would improve the usefulness of the tool in a business scenario. While the tool currently works exclusively on the level of process cohorts multiple participants thought it would be good to be able to *identify specific cases* in an event log (e.g. outliers) and to export those cases for further analysis using other tools. Furthermore, while people thought the visualisation enabled a quick comparison of cohorts they wanted additional features to *quantify these differences* and confirm their statistical significance. Participants also wanted to be able to *define more complex cohorts*, by splitting the data using multiple log attributes at once. It was suggested that the tool should be able to *guide the user to unexpected insights* in the analysis, rather than just confirm hypotheses the user has already got an intuition about. Finally, some participants were interested in seeing other perspectives such as the *resource and cost perspectives* integrated into the analysis. These suggestions are summarised in Table 6.

In summary, the participants liked the way data was visualised on the process model. However, the 3D aspect was received with mixed perceptions. The participants felt that the 3D is both appealing and made sense for the visualisation, but also felt that they would need training or at least more time to get used to it. On the other hand, perceptions of the tool's usefulness were more hesitant, suggesting that it would likely be of use once an appropriate target user group has been identified and trained in both process knowledge and the use of the tool. Furthermore, the limited availability of good process models and high quality process data could undermine the usefulness of the tool. These findings therefore reflect the sentiments present in the questionnaire results, but provide more in-depth insights into how these perceptions were formed.

5.3. Discussion

Overall, our evaluation provides support for the approach we have developed. While the user study uncovered several constraints that currently limit the usefulness of the proposed tool in a business environment, participants believed that organisations could use it to pinpoint the influence of different contextual factors on their complex processes quickly and easily, once issues with availability and quality of data are resolved and individuals received appropriate training. As the tool embodies our design principles, the evaluation of our tool in a broader sense also reflects on the utility of our design principles [44]. This also shows in some comments made by participants in the user study. People saying that the tool helped them to quickly and easily identify differences between cohort supports our design principle of presenting data in one integrated view for comparison (DP5). The comment that overview and drill-down features helped to identify areas on which to focus the analysis supports our principle of presenting data at different levels of abstraction (DP4). Furthermore, the suggestions that the visualisation integrates well with the stakeholders' mental model, could improve understanding of the process and would be useful for presenting analysis results indicate that the contextualisation of the data worked as intended, which provides support for our principle of using orthogonal dimensions to contextualise the performance data (DP6).

A factor that limits the validity of our conclusions regarding individual design principles is the question of how well our prototype

Table 5
Themes surrounding the usefulness of the tool.

Positive	Negative
• Could be used to draw insights, if constraints addressed • Would replace instinct with hard facts • Could improve understanding of process • Helped identify areas to focus analysis on • Good for managing complex processes • Useful for presentation of analysis findings	• Availability of good process model and data are perceived as limiting current analyses • Users need process and domain knowledge to interpret findings • Simpler tools exist for many use cases • Need to identify target audience: managers or process participants?

Table 6
Additional features suggested by interview participants.

• Should be able to identify and export cases from analysis
• Should have quantitative analyses and link numbers and visualisation more strongly
• Should have multidimensional splits
• Could have more dimensions
• Should be able to guide user to unexpected results

tool actually implements each principle, i.e. the instantiation validity of our tool [45]. As the principles abstract the design choices we have made for our approach, we expect that the tool implements all of the principles, however, the level of abstraction still leaves much flexibility in the design that could invalidate some of the conclusions. For example, observations that the arc visualisations can get crowded and that colour codes might affect the perceived dominance of certain cohorts in the visualisation can either indicate a problem of our implementation (i.e. our choice of visual glyphs and dimensions) or a problem with our principles to visualise all data in orthogonal dimensions and in one integrated view (DP5 & DP6). Overall, however, we regard the utility of the tool as an emergent property of the combined features captured in all the design principles put together. Therefore, while we generally regard the principles as falsifiable, we chose not to focus our evaluation on individual design principles. Consequently, we see the overall positive results of our evaluation as support for our approach in general and provisional support for the design principles.

While our evaluations provide some supporting evidence that the proposed approach can be of use to organisations there are some limitations to this claim. An important limitation stems from the lack of truly comparable approaches that also visually show performance differences for process cohorts. We chose not to perform a definitive and quantitative evaluation of our approach as this could only have been applied to aspects, such as the visualisation component, of the approach. The internal validity of our evaluation regarding the design principles is therefore limited. Instead we opted for increased ecological validity by using the approach with industry partners and applying it to industry datasets. This improved construct validity (with the exception of the link to the design principles) as we were able to measure utility in a realistic environment. However, we acknowledge that we only used subjective measures of this utility. In addition, the organisations used in the evaluation had fairly structured processes and were able to automatically log large parts of their processes. Consequently, both the representativeness and generalisability (i.e. the external validity) of our findings is somewhat limited as well and further research will need to establish whether organisations with more dynamic and difficult to log processes, for example in healthcare, would find the approach as useful. Furthermore, summarising statements into categories by paraphrasing carries some danger of researcher bias. We used dual-coding to reduce this risk, but interpreting natural language is inherently subjective. Thus, while we note these limitations may affect the level of support we found for our prototype tool, we believe that they lend at least tentative support for the main contribution of our work, which provides an approach for comparative process cohort analysis and the design principles that provide general guidance for the design of other process cohort comparison systems.

6. Conclusion

The interest in process mining is rapidly growing as is reflected by the number of commercial tools and real-life applications. Current tools provide functionality to analyse the performance and compliance of one process. Unfortunately, the comparison of cohorts of cases is seldom supported by contemporary tools. We argue that such comparisons are vital for process analysis. One may wish to compare different types of costumers, different outcomes, different process variants, different departments, etc. These process cohorts can be categorised based on any relevant context of interest (e.g. insurance claims submitted during storm season, patients with a high injury severity score that arrive at the hospital during peak hours). Consequently, this research proposed, implemented and evaluated a comparative process visualisation tool to address this issue. Firstly, requirements for such tool support were elaborated. A number of design principles are then proposed to meet these requirements. A prototype implementation of the approach as part of the process mining framework ProM demonstrated the viability of building a working system based on the proposed principles. Our validation of the tool with two industry datasets and practitioners shows that it can be applied to real world event log analysis scenarios and that business users see the tool as potentially useful. However, factors that currently limit the usefulness of the tool have also been identified. Future research therefore needs to focus on how to overcome these constraints and increase the perceived usefulness of the tool in organisations. In addition, the evaluation pointed to several ways the comparative capability supported by the tool can be extended. For instance, the tool can be extended to include additional perspectives in the visual comparison, such as the involvement of different resources/roles across multiple cohorts and the potential impact that could have on the performance statistics. Another interesting extension would be to add automated visual analytics algorithms to the tool to suggest points-of-interest for the user to investigate.

Acknowledgments

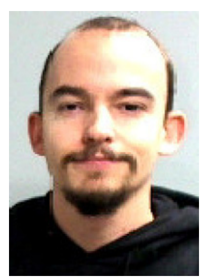
This work was supported by Australian Research Council (ARC) Discovery Grants DP120101624, DP150103356 and a 2014 Queensland Accelerate Partnership Grant (in partnership with QLD Motor Accident Insurance Commission). In addition we would like to gratefully acknowledge the contributions to this research made by the study participants from the Nominal Defendant (ND) and the Royal Automobile Club of Queensland (RACQ), as well as the input from our colleagues Robert Andrews and Willem Mertens.

References

[1] W.M.P. van der Aalst, Process Mining: Data Science in Action, Springer. 2016.
[2] S. Suriadi, M.T. Wynn, C. Ouyang, A.H.M. ter Hofstede, N.J. van Dijk, Understanding process behaviours in a large insurance company in Australia: a case study, in: C. Salinesi, M.C. Norrie, O. Pastor (Eds.), Advanced Information Systems Engineering, Lecture Notes in Computer Science vol. 7908, Springer. 2013, pp. 449–464.
[3] A. Partington, M. Wynn, S. Suriadi, C. Ouyang, J. Karnon, Process mining for clinical processes: a comparative analysis of four Australian hospitals, ACM Trans. Manag. Inf. Syst. 5 (4) (2015) 19:1–19:18.
[4] A. Bolt, M. de Leoni, W.M.P. van der Aalst, P. Gorissen, Exploiting process cubes, analytic workflows and process mining for business process reporting: a case study in education, International Symposium on Data-Driven Process Discovery and Analysis, CEUR-WS.org. 2015, pp. 33–47.
[5] A. Pini, R. Brown, M.T. Wynn, Process visualization techniques for multi-perspective process comparisons, in: J. Bae, S. Suriadi, L. Wen (Eds.), Asia Pacific Business Process Management, Lecture Notes in Business Information Processing vol. 219, Springer, Busan, Korea, 2015, pp. 183–197.
[6] W.M.P. van der Aalst, A. Adriansyah, B. van Dongen, Replaying history on process models for conformance checking and performance analysis, Data Min. Knowl. Disc. 2 (2) (2012) 182–192.
[7] S.J. Leemans, D. Fahland, W.M. van der Aalst, Exploring processes and deviations, International Conference on Business Process Management, Springer International Publishing. 2014, pp. 304–316.

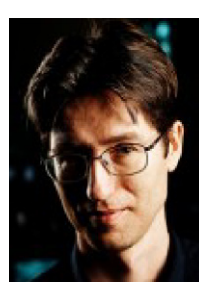
[8] M. de Leoni, M. Adams, W.M.P. van der Aalst, A.H.M. ter Hofstede, Visual support for work assignment in process-aware information systems: framework formalisation and implementation, Decis. Support. Syst. 54 (1) (2012) 345–361.
[9] B.F. van Dongen, A. Adriansyah, Process mining: fuzzy clustering and performance visualization, in: S. Rinderle-Ma, S. Sadiq, F. Leymann (Eds.), Business Process Management Workshops, Lecture Notes in Business Information Processing vol. 43, Springer. 2010, pp. 158–169.
[10] Fluxicon, Disco, 2016, https://fluxicon.com/disco/. version 1.9.3.
[11] A. Bolt, M. de Leoni, W.M.P. van der Aalst, A visual approach to spot statistically-significant differences in event logs based on process metrics, in: S. Nurcan, P. Soffer, M. Bajec, J. Eder (Eds.), Advanced Information Systems Engineering, Lecture Notes in Computer Science 9694, Springer. 2016, pp. 151–166.
[12] J. van Mourik, Comparing Business Processes Using Event Data: A Visual Approach, Technische Universiteit Eindhoven, Eindhoven, The Netherlands, 2015.Master's thesis.
[13] S. Kriglstein, G. Wallner, S. Rinderle-Ma, A visualization approach for difference analysis of process models and instance traffic, in: F. Daniel, J. Wang, B. Weber (Eds.), Business Process Management, Lecture Notes in Computer Science vol. 8094, Springer. 2013, pp. 219–226.
[14] J.C.A.M. Buijs, H.A. Reijers, Comparing business process variants using models and event logs, in: I. Bider, K. Gaaloul, J. Krogstie, S. Nurcan, H.A. Proper, R. Schmidt, P. Soffer (Eds.), Enterprise, Business-Process and Information Systems Modeling, Lecture Notes in Business Information Processing vol. 175, Springer. 2014, pp. 154–168.
[15] T. Mamaliga, Realizing a Process Cube Allowing for the Comparison of Event Data, Eindhoven University of Technology, Eindhoven, 2013.Master's thesis.
[16] T. Vogelgesang, H.-J. Appelrath, PMCube: a data-warehouse-based approach for multidimensional process mining, International Conference on Business Process Management, Springer. 2015, pp. 167–178.
[17] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C.D. Hansen, J.C. Roberts, Visual comparison for information visualization, Inf. Vis. 10 (4) (2011) 289–309.
[18] C. Healey, J. Enns, Attention and visual memory in visualization and computer graphics, IEEE Trans. Vis. Comput. Graph. 18 (7) (2012) 1170–1188.
[19] Minitlabs, Minit, 2016, http://www.minitlabs.com/. version 2.0.
[20] S. Malik, F. Du, M. Monroe, E. Onukwugha, C. Plaisant, B. Shneiderman, Cohort comparison of event sequences with balanced integration of visual analytics and statistics, Proceedings of the 20th International Conference on Intelligent User Interfaces - IUI '15, ACM, New York, New York, USA, 2015, pp. 38–49.
[21] R.C. Basole, M.L. Braunstein, V. Kumar, H. Park, M. Kahng, D.H.P. Chau, A. Tamersoy, D.A. Hirsh, N. Serban, J. Bost, et al. Understanding variations in pediatric asthma care processes in the emergency department using visual analytics, J. Am. Med. Inform. Assoc. 22 (2) (2015) 318–323.
[22] Z. Zhang, D. Gotz, A. Perer, Iterative cohort analysis and exploration, Inf. Vis. 14 (4) (2015) 289–307.
[23] K. Misue, S. Yazaki, Panoramic view for visual analysis of large-scale activity data, in: M. La Rosa, P. Soffer (Eds.), Business Process Management Workshops, Lecture Notes in Business Information Processing vol. 132, Springer. 2013, pp. 756–767.
[24] A.R. Hevner, S.T. March, J. Park, S. Ram, Design science in information systems research, MIS Q. 28 (1) (2004) 75–105.
[25] A.R. Hevner, A three cycle view of design science research, Scand. J. Inf. Syst. 19 (2) (2007) 4.
[26] T. Munzner, A nested model for visualization design and validation, IEEE Trans. Vis. Comput. Graph. 15 (6) (2009) 921–928.
[27] S. Gregor, A.R. Hevner, Positioning and presenting design science research for maximum impact, MIS Q. 37 (2) (2013) 337–355.
[28] A. Adriansyah, Performance Analysis of Business Processes From Event Logs and Given Process Models, Eindhoven University of Technology. 2009.Doctoral Thesis.
[29] B. Shneiderman, The eyes have it: a task by data type taxonomy for information visualizations, IEEE Symposium on Visual Languages, IEEE. 1996, pp. 336–343.
[30] D. Sacha, A. Stoffel, F. Stoffel, B.C. Kwon, G. Ellis, D.A. Keim, Knowledge generation model for visual analytics, IEEE Trans. Vis. Comput. Graph. 20 (12) (2014) 1604–1613.
[31] Z. Liu, J. Heer, The effects of interactive latency on exploratory visual analysis, IEEE Trans. Vis. Comput. Graph. 20 (12) (2014) 2122–2131.
[32] W.M.P. van der Aalst, The application of Petri nets to workflow management, J. Circuits Syst. Comput. 8 (01) (1998) 21–66.
[33] ProM Framework, 2016, http://www.promtools.org. version 6.6.
[34] J. Bertin, Semiology of Graphics: Diagrams, Networks, Maps, University of Wisconsin Press, Madison, Wisconsin, USA, 1983.
[35] D.L. Moody, The "physics" of notations: toward a scientific basis for constructing visual notations in software engineering, IEEE Trans. Softw. Eng. 35 (6) (2009) 756–779.
[36] W. Javed, N. Elmqvist, Exploring the design space of composite visualization, IEEE Pacific Visualization Symposium (Pacificvis), IEEE. 2012, pp. 1–8.
[37] W.P. Abbett, The magnetic connection between the convection zone and corona in the quiet sun, Astrophys. J. 665 (2) (2007) 1469–1488.
[38] M. Wattenberg, Arc diagrams: visualizing structure in strings, Proceedings of the IEEE Symposium on Information Visualization, IEEE, Los Alamitos, CA, USA, 2002, pp. 110–116.
[39] R. Baskerville, J. Pries-Heje, Explanatory design theory, Bus. Inf. Syst. Eng. 2 (5) (2010) 271–282.
[40] (DHBW) Karlsruhe, WoPeD, 2014, http://woped.dhbw-karlsruhe.de/woped/. version 3.2.0.
[41] S. Chaudhuri, U. Dayal, An overview of data warehousing and OLAP technology, ACM SIGMOD Rec. 26 (1) (1997) 65–74.
[42] F.D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology, MIS Q. (1989) 319–340.
[43] L.A. Palinkas, S.M. Horwitz, C.A. Green, J.P. Wisdom, N. Duan, K. Hoagwood, Purposeful sampling for qualitative data collection and analysis in mixed method implementation research, Adm. Policy Ment. Health Ment. Health Serv. Res. 42 (5) (2015) 533–544.
[44] J. Venable, J. Pries-Heje, R. Baskerville, A comprehensive framework for evaluation in design science research, International Conference on Design Science Research in Information Systems, Springer Berlin Heidelberg. 2012, pp. 423–438.
[45] R. Lukyanenko, J. Evermann, J. Parsons, Guidelines for establishing instantiation validity in IT artifacts: a survey of IS research, International Conference on Design Science Research in Information Systems, Springer. 2015, pp. 430–438.

Dr. Moe T. Wynn is a Senior Lecturer in the field of Business Process Management (BPM) within the School of Information Systems at Queensland University of Technology. Her research interests include process automation, process mining, comparative process analytics and cost-aware business process management. She has published over 60 refereed research papers. Her work appeared in well-known journals in the field including Information Systems, Information Sciences, Data and Knowledge Engineering, Information and Software technology, Formal Aspects of Computing, Journal of Computer and System Sciences, and Transactions on Petri Nets and Other Models of Concurrency.



Dr. Erik Poppe is a Research Fellow with the Business Process Management Discipline of the School of Information Systems at the Queensland University of Technology in Brisbane, Australia. With a background in the games industry and an interest in Information Systems, his research looks at improving technological support for human work practice through the application of game technologies and novel user interfaces. His current research focuses on the use of advanced visualisation techniques and visual analytics for comparative process analysis.

Dr. Jingxin Xu received his bachelor's degree in telecommunications engineering from Xidian University, Xi'an, China, in 2008, master's degree in information technology from the Queensland University of Technology (QUT), Brisbane, QLD, Australia, in 2010, and PhD degree in signal processing from QUT, in 2014. He is currently a Postdoctoral Research Fellow with QUT, investigating process mining techniques for business process management.



Prof. Arthur ter Hofstede is a Professor in the School of Information Systems in the Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia. He is also a Professor in the Information Systems Group of the Department of Industrial Engineering of the Technische Universiteit Eindhoven (TU/e). His research interests are in the areas of business process automation and process mining.

Dr. Ross Brown is a Senior Lecturer with the Science and Engineering Faculty at the Queensland University of Technology in Brisbane, Australia, where he is a member of the CHI Discipline, in the School of Electrical Engineering and Computer Science. Ross graduated with a PhD from QUT in 2003. He teaches computer graphics and final year project units in the Bachelor of Games and Interactive Entertainment. His main research interests are in the application of 3D games technology to other research domains. In particular, his latest research involves the development of virtual world technology for the representation of BPM information. A number of projects are currently underway, including: the embedding of executable workflows in virtual environments, collaborative 3D process modelling, 3D spatial visualisation of process model data and virtual world-based process elicitation. More information on his work can be found at: http://www.bpmve.org.

Dr. Azzurra Pini is a Research Fellow with the DensityDesign Research Lab (www.densitydesign.org) at Politecnico di Milano, Italy. Azzurra is a Communication Designer and her PhD research has been focusing on the role of visualization for the design of visual interfaces for exploring, analysing and designing organisational processes to assist researchers, analysts, and managers in their inquiry activity. Her current research interests pertain to all fields of application of communication and information design, including interdisciplinary research on complex phenomena and decision-support by means of visualization.

Prof.dr.ir. Wil van der Aalst is a Full Professor of Information Systems at the Technische Universiteit Eindhoven (TU/e), The Netherlands. He is also the Academic Supervisor of the International Laboratory of Process-Aware Information Systems of the National Research University, Higher School of Economics in Moscow. Moreover, since 2003 he has a part-time appointment at Queensland University of Technology (QUT). At TU/e he is the Scientific Director of the Data Science Centre Eindhoven (DSC/e). Wil van der Aalst has published more than 175 journal papers, 17 books (as author or Editor), 400 refereed conference/workshop publications, and 60 book chapters. Many of his papers are highly cited (he is one of the most cited computer scientists in the world and has an H-index of 126 according to Google Scholar) and his ideas have influenced researchers, software developers, and standardization committees working on process support. In 2012, he received the degree of doctor honoris causa from Hasselt University. In 2013, he was appointed as Distinguished University Professor of TU/e and was awarded an honorary guest professorship at Tsinghua University. He is also a member of the Royal Holland Society of Sciences and Humanities (Koninklijke Hollandsche Maatschappij der Wetenschappen) and the Academy of Europe (Academia Europaea).