



**HAL**  
open science

## Product Line Configuration Meets Process Mining

Houssem Chemingui, Ines Gam, Raul Mazo, Camille Salinesi, Henda Ben Ghézala

► **To cite this version:**

Houssem Chemingui, Ines Gam, Raul Mazo, Camille Salinesi, Henda Ben Ghézala. Product Line Configuration Meets Process Mining. CENTERIS, Oct 2019, Sousse, Tunisia. pp.199-210, 10.1016/j.procs.2019.12.173 . hal-02295954

**HAL Id: hal-02295954**

**<https://hal.science/hal-02295954v1>**

Submitted on 24 Sep 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies

## Product Line Configuration Meets Process Mining

Housseem Chemingui<sup>a,b,\*</sup>, Ines Gam<sup>a,b</sup>, Raúl Mazo<sup>a,c,d</sup>, Camille Salinesi<sup>a</sup>, Henda Ben Ghezala<sup>b</sup>

<sup>a</sup>Centre de Recherche en Informatique, Paris 1 Panthéon Sorbonne University, Paris, France

<sup>b</sup>RIADI Laboratory, National School for Computer Studies, Manouba, Tunisia

<sup>c</sup>Grupo de investigación GIDITIC, Universidad Eafit, Medellín, Colombia

<sup>d</sup>Lab-STICC, ENSTA Bretagne, Brest, France

---

### Abstract

Product line engineering is a new production paradigm that provides organizations a competitive edge by improving productivity and decreasing costs. The purpose with this new production paradigm is no longer to develop a single product but to develop a product family and to generate the products of the line through configuration processes. However, the potential benefits of product line engineering can be missed when dealing with large product lines because the configuration processes become error prone tasks. Consequently, guiding stakeholders during such complex configuration processes and recommending the best configuration alternatives until leading to a satisfying experience becomes a challenge. This paper focuses on enhancing product line configuration processes through process mining techniques. Therefore, user's actions of previous product line configurations are logged, mined and analyzed. We conducted a preliminary research to motivate the advantages of process mining in product line configuration and to explore what can process mining bring for configuration processes and how to use it to enhance configuration processes. Thus, guidance questions are sketched in order to position process mining as a solving tool for the configuration difficulties. Furthermore, we propose a reference architecture that considers process mining for configuring product lines.

© 2019 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the scientific committee of the CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies

*Keywords:* Product line engineering; configuration process; process mining, enhancing; configuration difficulties.

---

---

\* Corresponding author.

*E-mail address:* [mohamed-housseem.chemingui@etu.univ-paris1.fr](mailto:mohamed-housseem.chemingui@etu.univ-paris1.fr)

## 1. Introduction

Product Line Engineering (PLE) is a reuse-driven production paradigm that has been successfully applied in several kinds of industries, such as software [28], automotive [29][30], avionics [31] and electronics [32]. Extensive research and industrial experience have widely proven the significant benefits of PLE practices, among them are: increased reuse, productivity and quality, and reduced time to market [28]. These benefits are obtained through the application of two main processes called domain engineering and application engineering. The domain engineering process is designed to capitalize and document all artefacts necessary to produce systems; and the application engineering involves the configuration process that constitutes the activities of deriving a valid product from the product line [33].

In this sense, configuration activities should meet the requirements of the user and respect the constraints of the Product Line Model (PLM). Usually, when the number of features increases in a PLM, the number of potential products increases also. Indeed, the configuration process becomes quickly complex when the number of user decisions to be made increases; the fact that these decisions must be taken in a predefined order that respects user preferences on the one hand and the PLM constraints on the other hand. Therefore, user assistance becomes central especially when users are dealing with many possible combinations of configuration decisions.

Throughout unguided configurations, the PLM is configured without following a sequence of well defined choices. It consists of making decisions by browsing the PLM randomly. Staged configuration processes [34] are typical examples of non-guided configurations. Different approaches were proposed to direct users towards the configuration solutions corresponding to their expectations such as the configurations driven by constraint satisfaction problems [35] and the configurations using the recommendation [18]. Nonetheless, these proposals present drawbacks that will be discussed in this paper. Hence, the idea is to consider process mining [36] as a guidance strategy that considers previous process behaviors to anticipate the behaviour of future configuration processes.

Consequently, the goal of this paper is to explore and discuss the contribution of process mining in overcoming the shortcomings of product line configurations. Based on the problem description, this research was supported by a Design Science Research (DSR) [37] as the goal is to expose problems, suggest solutions and motivate their use. For the execution of a typical DSR methodology, this research was guided by the following steps [38]: (i) Identifying the problems of the configuration process; (ii) suggesting a proposal that handles the encountered problems; (iii) proposing an architecture that support solutions; (iv) comparing the proposal to other existent solutions; and (v) highlighting the additional knowledge with regards to the acquired process acquired. Thus, this research aims to propose a qualitative approach with an exploratory nature that focuses on what process mining can bring to product line configurations. Other domain engineering processes such as the formalization of PLMs [35][39] and their verification [40][41][42] are not taken into account. This purpose leads to the following research questions:

**RQ1:** What are the problems that users meet when they are configuring product line models?

**RQ2:** What challenges and guidance questions can be inferred from process mining to assist the configuration process?

**RQ3:** Which technical solution to support the guidance challenges and what process traces are useful to guide the configuration of product line models?

**RQ4:** How useful to adopt a process mining strategy to enhance the product line configuration?

For this purpose, our studies rely on some recent publications in the field. The main concern is to make a preliminary analysis of the needs of product line configurations. Thus, this study is neither exhaustive nor systematic. It presents a continuity of our research works on product line engineering solutions [17] [43] and process mining investigations [44][45][22][46]. Besides, this paper is a part of a larger scientific project aiming to propose recommendation methods for product line configuration [17][18].

The rest of the paper is organized as follows. Section 2 outlines configuration problems and the guidance challenges aiming to address them. Section 3, introduces the process mining concepts and proposes a process mining based architecture that supports the expected solutions. Section 4 explores the usage outcomes of process mining as a problem-solving tool for product line configuration issues. Section 5 discusses the related works and the current approach limits while Section 6 concludes the paper and discusses research perspectives.

## 2. Configuration Problems and Guidance Challenges [RQ1 & RQ2]

To answer **RQ1**, we conducted a literature review on recent papers published in journals and conference proceedings that treated the subject of product line configuration issues. A key element in literature reviews is the explicit definition of a review process that guides its execution. The main stages of our review process are :

- The definition of research question
- The definition of search string and digital libraries
- The selection and qualification of papers
- The data extraction and synthesis

This study is not systematic in the sense of [65] or [66]; however, it follows the recommendations of Webster and Watson [53] to conduct literature reviews. According to our research question, we conducted a literature review rather than a mapping study. This choice was motivated by the fact that we aim to identify, analyze and interpret evidence related to a specific research question and not a topic of area.

We referred in the search string to the title and the abstract of the paper. The authors collectively elaborated the reference lists of the most important journals, conferences and other venues. The review included literature published from 2006 to 2018 reporting on research issues for product line configurations. Using Google Scholar and the free search feature of DBLP, the following search terms concerning product line and difficulties were used and combined: “product line”, “product line engineering”, “product family engineering”, “variability”, “problems”, “issues”, “difficulties” and “shortcomings”. As the total number of retrieved publications was limited, we used Google Scholar to browse publications that cited previously select papers, i.e., forward snowballing . Using this technique, further papers were added to our list.

We conducted the literature review in 28 relevant sources. The following exclusion criteria were defined in the selection stage in order to narrow the current list of candidate papers:

- Criteria 1 : We excluded papers that were not available in electronic form.
- Criteria 2: Only publications written in English and french were kept. For example, papers written in Spanish and german were excluded.

Table 1 focuses on the encountered major issues of product line configurations and in which references they were mentioned. These configuration issues are classified into the following three categories :

- *Decision complexity*: this category summarize some technical difficulties that users are facing when they deal with unguided configurations.
- *Lack of performance*: this category lists problems, encountered by existent configuration process executions, in relation with computation time and the heavy charge including unnecessary evaluations.
- *Enhancement need*: a trend in recent research is to go a step further, beyond the explicit request of relevant guidance solutions in product line configuration processes.

Table 1: Major issues of product line configurations

Issues Focus	References
<b><i>THE DECISION COMPLEXITY</i></b>	
Combinatorial explosion of variants causing errors and conflicting decisions: heavy human effort.	[2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [14], [15], [16], [17], [19], [20], [21], [22], [23], [24], [25],[57]
Configuration Quality: efficiency, completeness and flexibility.	[1], [2], [3], [4], [6], [8], [9], [10], [15], [19], [20], [21], [23], [24], [26], [27]
Need to reveal the impact of decisions.	[9], [17], [18], [25]
<b><i>THE LACK OF PERFORMANCE</i></b>	
Undoing and repeating.	[13], [15], [17]
Lengthy delivery time.	[1], [2], [3], [4], [6], [10], [11], [13], [14], [15], [17], [18], [24]
<b><i>THE ENHANCEMENT NEED</i></b>	

Lack of user support.	[7], [9], [10], [11], [14], [16], [17], [18], [22], [24], [26], [27]
Prediction and recommendation need.	[7], [12], [17], [18], [19], [27]
Decision automation need.	[2], [5], [14], [17]
Visual and interactive support need.	[7], [8], [9], [18], [20], [26]
Limited insight and information of decision impacts.	[9], [18], [25]
Lack of personalization.	[1], [7], [15], [18], [22]
Need to elicit user requirements, preferences and goals.	[2], [6], [11], [12]
Need to map the elicited configuration goals and the PLM elements.	[7], [10], [11], [12], [24]

Table 1 shows that 80% of selected studies mentioned the problem of *variant explosion causing error prone processes and conflicting decisions* during the configuration process. Large PLMs can have several hundred features and the solution space can have several thousands or more constituents. Thus, it becomes hard to understand the multiple decisions to make. The second most mentioned issue concerns also the decisions complexity and is about the *configuration quality* including: *efficiency, completeness and flexibility*. The users are all time facing PLMs with no efficient configurations. Therefore, users leave the process before finishing all the decisions to make since the provided choices are, also, not flexible. The third issue, concerning the decisions complexity, is the *need to reveal the impact of decisions* since the configuration is based on PLM containing constraining variants. At each configuration stage, the choices can cause the automatic propagation of selection or rejection of other variants. Thus, it is crucial to prevent the user of the impacts of his actions before being faced to a reduced solution space that don't exactly meet to his intentions and requirements. The performance of the configuration process was also a major problem category found in the literature review. Indeed, users are often faced to *undo choices and repeating cycles* causing backtracks. Several research works revealed that *long time delivery* of valid combinations is a significant impediment to the configuration process. The users need, in most cases, valid configurations with fast feedbacks resulting in lower lead times. This problem is frequent especially in large product lines that need complex computations in order to reach a valid configuration. These researchers find that is essential to think about elastic solvers which can scale up regardless the PLM size.

Other researches have gone further to bring the need to enhance product line configuration process. Indeed, the majority of analyzed papers mentioned the *lack of user support* provided to users when they are configuring a product. To tackle these difficulties, some researches mentioned the need of using *prediction and recommendation* techniques during the configuration process (e.g., recommend values of features according to predictions based on similarity measures). Other works mentioned the need of *automatic decisions* especially in situations when there is no choices. Another manner to enhance the configuration process, it to provide to users ergonomic interfaces with *visual and interactive support*. To reveal the impact of decisions some studies bring the need to give users *insights and messages about their action consequences*. The *lack of personalization* in provided choices to the users, is mentioned also as a major issue since it constitutes the aim of the product line method. Mentioned research works preconise that to enhance the configuration process it is crucial to *elicitate requirements, preferences and goals* of each users. Besides, five works highlight the need of *mapping the elicited requirements, preferences and goals with the assets of the corresponding PLM* (e.g., the requirements, constraints and the domain components of the PLM).

Regarding **RQ2**, the expectations of users who are configuring a product line are hard to be precisely predicted and represented in advance. The enactment of a single configuration process that works on all the cases is impractical since the configuration combinations are heterogeneous and complex. Furthermore, providing to users error prone and annoying processes might not be very beneficial or profitable. Moreover, the measurement of process performance is a natural part of the configuration process development. All this might raise questions about efficiency and effectivity related to process improvement. For this reason, guidance solutions based on process mining techniques can provide

an added value. Thus, this research outlines the following four major guidance challenges for the product line configuration processes and their corresponding guidance questions.

Challenge 1: Provide flexibility during the configuration process.

Challenge 2: Maximize the personalization of the configuration suggestions.

Challenge 3: Reduce the required time to reach a suitable configuration.

Challenge 4: Promote configurations that maximise economic benefits.

For each challenge, guidance questions aiming to inspect the configuration issues are listed.

#### *Guidance questions for Challenge 1*

GQ 1.1: What are the safest (doubtful) processes that lead to a valid (incoherent) product?

GQ 1.2: What are the processes that include fashionable products?

GQ 1.3: What are the processes that include the best/worst seller products?

GQ 1.4: What are the processes that include full/minimum option products?

GQ 1.5: What are the processes that were interrupted by the system or by the user?

GQ 1.6: What are the processes that were configured by serious users? (not just trying combinations)

GQ 1.7: What are the processes that are executed with a maximum number of manual (user) choices?

GQ 1.8: What are the processes that are executed with a maximum number of automatic propagations?

#### *Guidance questions for Challenge 2*

GQ 2.1: What are the processes of a specific user based on his history?

GQ 2.2: Which users have similar configuration intentions?

GQ 2.3: What are the processes that include the best/worst seller products for each category of users (e.g., gender, age, country) or a specific period (e.g., summer holidays)?

#### *Guidance questions for Challenge 3*

GQ 3.1: What should be configured first?

GQ 3.2: What is the average number of steps or required time to execute a process?

GQ 3.3: What is the ideal process order that reduce the solver charge?

GQ 3.4: What are the processes that were executed rapidly/slowly?

#### *Guidance questions for Challenge 4*

GQ 4.1: What are the processes that correspond to products that minimise the cost?

GQ 4.2: What are the processes that correspond to products that maximise the profit?

GQ 4.3: What are the processes that correspond to products with less human resources?

GQ 4.4: What are the processes that lead to sell accumulated stock?

Confronting the literature review study with guidance challenges, we denote that challenges 1 and 2 handle the decision complexity and the enhancement need issues. Nonetheless, challenge 3 handles the lack of performance in existent configurations processes. Challenge 4 aims to guide users by suggesting to them choices that promote business interests of the product line provider.

On the other hand, studies over the last few years have associated some benefits that can be achieved by using process mining for complex business issues. These benefits were analyzed on a survey proposed by Claes et al. [54]. Revealed benefits are in correspondence with the guidance challenges and configuration issues. However, this alignment is neither systematic nor exclusive and defining a combination of challenges and issues at each guidance situation lead to better product line configurations. Interpreting process mining findings, based on traces and generated models, is useful to answer the evoked guidance questions and to provide assistance during the configuration process.

To support revealed challenges of product line configurations, we shape in the next section a reference architecture based on process mining.

### **3. Process mining at the service of product line configuration [RQ3]**

Process mining was defined by Günther [47] as an area of research that is concerned with posteriori analysis of business processes, based on event logs of their execution. Its purpose is to extract global and specific information and highlight the different aspects of these business processes. Hence, the process mining is adopted by some organizations that need to have enough information about what is happening in their activities.

Moreover, process mining is a set of techniques that enable to discover, monitor and improve processes through the extraction of information from event logs available in information systems [36]. The starting point is an

information system that records executions' traces of processes and store them in an event log. An event log can be seen as a collection of cases and a case as a sequence of events. Each event contains information about its execution, such as the timestamp, the activity that involve, the originator of this event, etc. According to these traces, process models allow to show execution details generated thanks to techniques of process mining and especially the following three ones: (i) Discovery : to mine process models ; (ii) Conformance : to identify, explain and quantify process differences when existing process models are compared with the discovered ones; and (iii) Enhancement : to apply changes to the existing process models based on the discovered ones. Furthermore, some techniques aim to go further and use traces to suggest which tasks may follow a current activity [48]. The process mining techniques are, mainly, process-driven implying the discovery of rich, high-level models with decision points, loops and concurrency [45]. Moreover, the process models are complete, being able to capture behaviors which are also less frequent in the data mining techniques, such as sequence mining [49] and episode mining [50].

In this research, we focus on process discovery and enhancement. To do so, we analyse configuration frequency (e.g., activities and paths frequencies), configuration performance (e.g., duration and computation costs), and the similarity measures between processes.

To answer **RQ3**, we conducted a proof of concept by using a collection of configured products from PLMs where the different users are not aware of what aspects are going to be studied. Preliminary experiments were held in order to evaluate what process mining can bring to support the guidance challenges. Moreover, experiments were iterative and they allowed us to improve the drawn reference architecture.

This research studies the ways in which users configure products from a product line with the aim of guiding them in future configuration. In particular, this research explores the use of process mining techniques to guide users during future configuration processes. With this technique, configuration traces of previous configuration processes were extracted and saved into a trace repository. Traces were collected with the VariaMos<sup>†</sup> [51] product line management tool and process mining analysis of these traces were done with Disco<sup>‡</sup> [52]. Disco deal with event logs to discover process models that can reflect configuration process issues. The exact process mining algorithms used in Disco are not public but the owners claim that a technique based on Fuzzy Miner [58] is implemented. Fuzzy Miner is aimed at exploratory analysis, in particular, to discover dependency graphs, which are directed graphs with activities as nodes and process relations as edges. Dependency graphs discovered with the fuzzy miner algorithm are able to capture complex relations, such as concurrency, loops and indirect sequences. Moreover, to filter out the log noise, multiple dependency measures are defined, customized for specific types of relations. Filters allow to generate and interpret process models in different ways. In this way, the process representation varies from showing all observed patterns in the log to showing only the most significant patterns. Process models representation and interpretation are on correspondence with the guidance questions listed for each challenge.

To do so, a reference architecture was designed in a layered composition including a presentation, a mining and a recommendation layers. This architecture is based on four main core modules: the *Configurator* at the presentation layer, the *Event Logs Collector* and the *Process Miner* at the mining layer and the *Process Recommender* at the recommendation layer. Figure 1 shows this reference architecture where bold police mark these modules. Indeed, the Configurator proposes a graphical interface that supports the configuration process and configuration decisions for a given product line model. The Event Logs Collector records filtered event logs issued from the configuration processes; and the logs processing allows recording these event logs according to a specific structure. Event logs structure is designed in a way that allows capturing information concerning the process; the elements of the product line model involved in each configuration event, the event order and the time stamp. Considering process mining techniques and guidance questions listed above, event logs structure is defined as shown in Table 2.

---

<sup>†</sup> <https://variamos.com/>

<sup>‡</sup> <https://fluxicon.com/disco/>

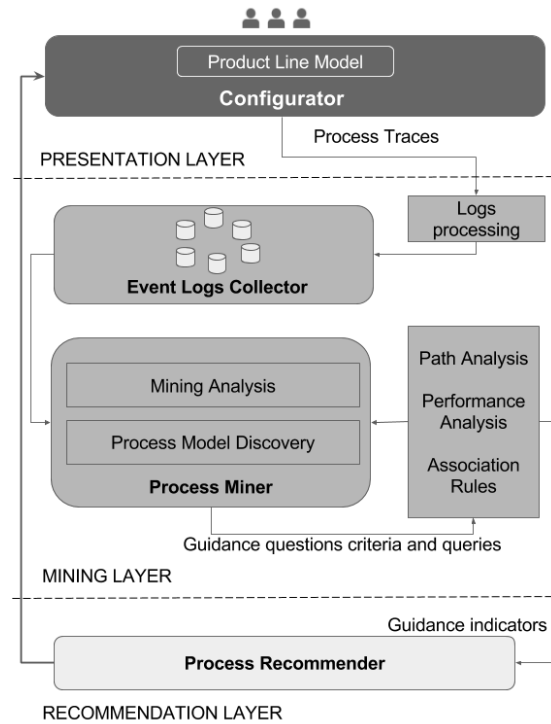


Figure 1- Reference architecture for mining product line configuration processes.

The process miner encapsulates two main activities namely Mining Analysis and Process Model Discovery. In order to answer guidance questions, the definition of mining criteria and queries is important. Therefore, we analyse path executions (activities transitions), performance results and association rules between configuration process fragments. The choice and the settlement of process mining algorithms depend on the required guidance goals. Process mining interpretations can be, then, done according to arised guidance indicators.

The Process recommender module takes as input the indicators from process mining interpretations and provides recommendations to the Configurator module, which is in charge of proposing these recommendations to the final users in an ergonomic way.

Table 2: Definition of the event logs structure

Attribute	Description
CP ID	The configuration process identifier
Originator	The user identifier
PLM element ID	The configured element (e.g., feature, domain asset) identifier
PLM element type	The type of the configured element (e.g., mandatory,, grouped or optional)*
PLM element value	The value that the user assigned for the given product line element (e.g., 1 when the corresponding element is selected and 0 when it is rejected)
Decision type	The type of configuration decision (manual or automatic propagation of choices)
Decision order	The order of the decision in the process
Timestamp	The timestamp of the decision

\* The type of the configured element depends on the modeling notation: OVM, Feature Model, etc.

The proposed reference architecture aims to mine product line configuration processes. The main concern of this technical solution is to meet the process mining outcomes and prove what they can bring for the configuration



shortcomings. Consequently, the choice of a layered architecture was motivated by the fact that several functionalities must be included in the technical solution, in particular: *presentation*, *mining* and *recommendation*. This process mining based architecture aims to propose an additional layer of process mining, compared with other architectures of software configurators based largely on the classical recommendation paradigms [17] or on the simple propagation of domain constraints [67]. The mining layer is considered as the essence of the recommendation layer. Adopting a process mining strategy aims to select dynamically variants according to the progress of the configuration process. In instance, the order of the process steps and the variants values are not standard; they differ according to the user goal and the partial configuration.

#### 4. Process mining outcomes for Product Line configuration processes [RQ4]

To answer **RQ4**, we conducted a literature search on process mining outcomes. As a matter of fact, process mining has helped in some of business process issues. According to our basic experiments on mining product line configurations, we selected, adapted and extended the outcomes of process mining [54][55] that are strongly coupled and useful for the decision complexity, the lack of performance and the need of enhancement discussed in Table 2.

##### *Outcome 1: Objectivity and transparency*

Process traces reflect the manner of how a given configuration process has exactly been executed. Thus, process models reconstructed through different process mining techniques are not subject to perceptual biases. Process mining enables to “drill-down to the facts” to obtain information about who selected/rejected a variant from the product line, how much time did it take, etc. Other frequency analysis can be held to discover, for example, what is the path in the configuration process that is followed by the highest percentage of cases.

##### *Outcome 2: Process speed-up*

Though a first basic understanding of the business upfront, it is obvious that reproducing configuration process diagrams may considerably accelerate future configuration activities. In fact, logging the timestamp of each user action allows identifying the mean, minimum and maximum duration of a configuration process.

##### *Outcome 3: Reduced effort (efficiency)*

Process discovery enables a considerable increase of efficiency thanks to the fact that less effort to discover business issues is needed. In an enhancement context of the configuration process, based on previous traces, it becomes possible to recommend to users choice points that avoid following branches containing slowly activities.

##### *Outcome 4: Root and bottlenecks*

Process mining can be used to easily find out the root causes of configuration difficulties, bottlenecks and undo choices that are frequent in product line configurations. Undoing variable values can affect other variable values, which obviously causes real calculating problems for the solver in charge of the product line configuration processes support.

##### *Outcome 5: Prediction and guidance*

Based on facts, it is possible to predict future configuration process executions. Thus, event logs can be used to propose more reliable and guided configuration processes than the ones purely based on theoretical assumptions. This outcome is the main concern of this paper in front of the enhancement need revealed in literature and industry.

All guidance questions are important to provide guidance during product line configuration processes. Process mining analysis were able to provide answers for these guidance question by focusing on path frequency, time performance, originator tasks, sequence orders, undo cycles, etc. However, guidance questions that correspond to challenge 4 need further business information about stock and costs of configured products.

#### 5. Related works

Difficulties of the software configuration process have been addressed by the product line community that proposed interactive configuration methods [59][60], recommendation-based methods [61], combination of these two methods [18] and heuristic-based methods [17].

In interactive methods, product line specifications are mapped into an executable language, for instance, finite domain constraint programs [35] and propositional formulas [62]. The goal of this transformation is to automatically propagate user decisions and provide a valid solution when there is only one possible choice. Our work with companies [29] revealed that this is very efficient to help stakeholders to identify inconsistent decisions, but that the probability of inconsistent decisions was so high that this approach is virtually impossible to use when stakeholders don't already have a deep expert knowledge of the systems to configure.

In recommendation-based methods, configurations are mainly based on "content-based" filtering [63], and/or "collaborative" filtering [64]. For instance, [19] uses data mining to recover complex feature correlations and association rules in order to make recommendations. Other works, such as [18] propose an interactive configuration by organizing the configuration process in a series of stages where decisions are made using recommendations based on the similarity between former decisions and all other decision processes that were made before. This approach has 2 important drawbacks. First, it suffers of the cold-start issue, which is typical of all collaborative filtering strategies. Second, it creates cognitive reinforcement : stakeholders receive recommendations that correspond to their way of working, rather than suggesting other potentially more adequate or efficient approach.

In heuristic-based methods, users are guided throughout the configuration by using heuristics that aim to reduce the number of user actions and/or minimize impact of decisions [17]. Heuristics are intended to help customers to specify the characteristics of their products step-by-step according to their requirements. The main interest of this research is to improve the configuration performance by avoiding inefficient decisions., however, other configuration issues such as the personnalisation of guidance seem conscient but not overcome. This approach does not recommend a process but it recommends features of a product using heuristics and not using similarity on past configurations.

Other methods, propose to deal with complex configurations and provide intelligent configuration processes, the notions of views and workflows have been defined [33]. This kind of assistance is provided especially to handle collaborative configurations that involve different users with different concerns. Nevertheless, this method presents limits since it don't consider user intentions and don't focus on process execution behaviors. In fact, proposed guidance solutions are standard; they don't adapt to the variety of support needs and situations faced by the configuration collaborators. Moreover, it provides limited choices that make users face inflexible processes; that may suggest variants values but never a configuration process path and order.

As far as we know, there are no studies that explicitly attempt to apply process mining on this context. Whereas, many other works tried to capture configuration traces and exploit them in order to enhance the product line configuration. In this paper, this kind of research is considered as an implicit proposal of configurations mining. In this sense, [27] proposed to suggest to users the next action when they are building a configuration. The authors used the result of artificial intelligence clustering to suggest actions. They capture action traces and classify them by using the clustering measures and determine the closest cluster. However, this research also provides to users suggestion about an only action to make and not a series of actions to follow. Adding to that, the suggested action is based on its frequency in traces and don't take into account any guidance goal.

Revealed limits motivated the idea of adopting a process mining strategy that aims to analyze the whole process executions. The purpose is to guide future configurations by suggesting, at each stage, values for a given PLM element on the one hand, and configuration paths or sequence of activities, in a predefined order, in the other hand. All guidance solutions take into account a guidance goal, a process state and a given situation.

## 6. Conclusions and Perspectives

A guided enactment of the configuration process, still faces difficulties to emerge from the plethora of possible combinations, differents situations and heterogenous guidance goals. For a deep understanding of the configuration process difficulties, an exploratory study was held and revealed the following main problems: The decision complexity, the lack of performance and the numerous enhancement needs (**RQ1**). As an answer, this paper proposes to mine previous configuration processes, discover process models and interpret them to improve future processes. This process mining strategy was motivated by the following three challenges: The decision flexibility and personalization, the configuration performance and the economic interests consideration (**RQ2**). To meet these challenges, a set of guidance questions is outlined to meaningfully conduct the interpretation of discovered process models. To do so, a reference architecture was designed in a layered composition including a presentation, a mining and a recommendation layers (**RQ3**). The potential outcomes of our preliminary experimentations showed an objectivity and transparency of discovered process models, their ability to speed-up processes, providing efficiency, identifying process bottlenecks and guiding next processes (**RQ4**). Mining product line configuration processes can go a step further than suggesting a value for a configuration option, at a given stage of configuration, and suggest a serie of actions of the whole process. This analysis based on processes (and not static data), can reveal informations about the execution behavior. Thus, suggesting to users a configuration order and variants values according to a guidance goal proves original compared with existing recommendation techniques in software configurators.

Nonetheless, several scientific questions are arisen concerning the consequences and drawbacks of this proposal for practitioners. When the PLM evolve, guidance directives meet a significant gap between what is capitalized in traces and what the new evolved PLM contain. Consequently, the guidance based on the previous PLM becomes incoherent with the new PLM configuration consistencies and lead to the following questions: What artefacts can be exploited and reused in spite of new changes? Even further beyond, what dependencies and correlations exist between the configuration process and the PLM? What is the dependence degree between a configuration process and the PLM? Furthermore, the cold start phenomenon has to be well controlled when process mining guidance is intended. In the beginning, trace-based analysis fail to draw sensible guidance directive due to the lack of informations about past user interactions. Thus, in trace-dependent recommendation, since more traces are extracted, the efficiency and accuracy of guidance are reinforced. Last but not least, mining domain engineering processes constitutes a perspective of this work also. The usefulness of process mining in verifying the consistency of a PLM or identifying its bottlenecks and their causes seems a promising research axis. Future work will focus also, among others, on the event logs extraction with capturing informations including the stakeholders profiles. Moreover, automatic identifying of the stakeholders goals will be tackled more thoroughly by a deeper focus on the GQM method [56].

## Acknowledgements

This work is supported by the French–Tunisian CMCU project 16G/1416 called CONFIGURE.

## References

- [1] Horcas, J. M., Pinto, M., & Fuentes, L. (2017, September). Green configurations of functional quality attributes. In *Proceedings of the 21st International Systems and Software Product Line Conference-Volume A* (pp. 79-83). ACM.
- [2] Noorian, M., Bagheri, E., & Du, W. (2017). Toward automated quality-centric product line configuration using intentional variability. *Journal of Software: Evolution and Process*, 29(9), e1870.
- [3] Ochoa, L., Pereira, J. A., González-Rojas, O., Castro, H., & Saake, G. (2017, February). A survey on scalability and performance concerns in extended product lines configuration. In *Proceedings of the Eleventh International Workshop on Variability Modelling of Software-intensive Systems* (pp. 5-12). ACM.
- [4] Hayashi, K., Aoyama, M., & Kobata, K. (2017, September). Agile Tames Product Line Variability: An Agile Development Method for Multiple Product Lines of Automotive Software Systems. In *Proceedings of the 21st International Systems and Software Product Line Conference-Volume A* (pp. 180-189). ACM.
- [5] Zhan Z., Zhan Y., Huang M., Liu Y. (2018) Product Configuration based on Feature Model. In: Xhafa F., Patnaik S., Zomaya A. (eds) *Advances in Intelligent Systems and Interactive Applications. IISA 2017. Advances in Intelligent Systems and Computing*, vol 686. Springer, Cham.
- [6] Afzal, U., Mahmood, T., & Shaikh, Z. (2016). Intelligent software product line configurations: A literature review. *Computer Standards & Interfaces*, 48, 30-48.
- [7] Pereira, J. A., Matuszyk, P., Krieter, S., Spiliopoulou, M., & Saake, G. (2017). A feature-based personalized recommender system for product-line configuration. *ACM SIGPLAN Notices*, 52(3), 120-131.
- [8] Lu, H., Yue, T., Ali, S., & Zhang, L. (2016, April). Nonconformity resolving recommendations for product line configuration. In *2016 IEEE International Conference on Software Testing, Verification and Validation (ICST)* (pp. 57-68). IEEE.
- [9] Sion, L., Van Landuyt, D., Joosen, W., & de Jong, G. (2016, September). Systematic quality trade-off support in the software product-line configuration process. In *Proceedings of the 20th International Systems and Software Product Line Conference* (pp. 164-173). ACM..
- [10] Sinnhofer, A. D., Pühringer, P., Potzmader, K., Orthacker, C., Steger, C., & Kreiner, C. J. (2016, June). A framework for process driven software configuration. In *BMSD 2016* (pp. 1-8).
- [11] Bagheri, E., & Ensan, F. (2014). Dynamic decision models for staged software product line configuration. *Requirements Engineering*, 19(2), 187-212.
- [12] Noorian, M., Bagheri, E., & Du, W. (2014). From Intentions to Decisions: Understanding Stakeholders' Objectives in Software Product Line Configuration. In *SEKE* (pp. 671-677).
- [13] Sayyad, A. S., Ingram, J., Menzies, T., & Ammar, H. (2013, May). Optimum feature selection in software product lines: Let your model and values guide your search. In *Proceedings of the 1st International Workshop on Combining Modelling and Search-Based Software Engineering* (pp. 22-27). IEEE Press.
- [14] Guana, V., & Correal, D. (2013). Improving software product line configuration: A quality attribute-driven approach. *Information and Software Technology*, 55(3), 541-562.
- [15] Elsner, C., Ulbrich, P., Lohmann, D., & Schröder-Preikschat, W. (2010, September). Consistent product line configuration across file type and product line boundaries. In *International Conference on Software Product Lines* (pp. 181-195). Springer, Berlin, Heidelberg.
- [16] Mendonça, M., Bartolomei, T. T., & Cowan, D. (2008, March). Decision-making coordination in collaborative product configuration. In *Proceedings of the 2008 ACM symposium on Applied computing* (pp. 108-113). ACM.

- [17] Mazo, R., Dumitrescu, C., Salinesi, C., & Diaz, D. (2014). Recommendation heuristics for improving product line configuration processes. In *Recommendation Systems in Software Engineering* (pp. 511-537). Springer, Berlin, Heidelberg.
- [18] Triki, R., Mazo, R., & Salinesi, C. (2014). Combining configuration and recommendation to enable an interactive guidance of product line configuration.
- [19] Zhang, B. (2013). Mining Complex Feature Correlations from Large Software Product Line Configurations.
- [20] Botterweck, G., Nestor, D., PreuBner, A., Cawley, C., & Thiel, S. (2007). Towards supporting feature configuration by interactive visualisation.
- [21] Zhang, G., Ye, H., & Lin, Y. (2010, November). Quality attributes assessment for feature-based product configuration in software product line. In *2010 Asia Pacific Software Engineering Conference* (pp. 137-146). IEEE.
- [22] Deneckère, R., & Kornysheva, E. (2010). Process line configuration: An indicator-based guidance of the intentional model map. In *Enterprise, Business-Process and Information Systems Modeling* (pp. 327-339). Springer, Berlin, Heidelberg.
- [23] Guo, L. (2006). Consumption flexibility, product configuration, and market competition. *Marketing Science*, 25(2), 116-130.
- [24] Dumitrescu, C., Tessier, P., Salinesi, C., Gérard, S., & Dauron, A. (2013). Flexible product line derivation applied to a model based systems engineering process. In *Complex Systems Design & Management* (pp. 227-239). Springer, Berlin, Heidelberg.
- [25] Mendonca, M., Cowan, D., & Oliveira, T. (2007, January). A process-centric approach for coordinating product configuration decisions. In *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)* (pp. 283a-283a). IEEE.
- [26] Urli, S., Perez, G., Zitoun, H., Blay-Fornarino, M., Collet, P., & Renevier-Gonin, P. (2012, November). Vers des interfaces graphiques flexibles de configuration. In *Journée Lignes de Produits 2012*.
- [27] Chastel, F., Blay-Fornarino, M., Pereira, C. D. C., & Urli, S. (2015). Apprendre des actions utilisateur pour guider le processus de configuration dans les lignes de produits logiciels.
- [28] Clements, P., Northrop, L. *Software Product Lines: Practices and Patterns*, Addison-Wesley, 2001. doi:10.1111/j.1365-2648.2011.05843.x.
- [29] Dumitrescu, C., Mazo, R., Salinesi, C., & Dauron, A. (2013, August). Bridging the gap between product lines and systems engineering: an experience in variability management for automotive model based systems engineering. In *Proceedings of the 17th International Software Product Line Conference* (pp. 254-263). ACM.
- [30] Young, B., Cheatwood, J., Peterson, T., Flores, R., & Clements, P. (2017, September). Product Line Engineering Meets Model Based Engineering in the Defense and Automotive Industries. In *Proceedings of the 21st International Systems and Software Product Line Conference-Volume A* (pp. 175-179). ACM.
- [31] Sharp, D. C. (2000). Component-based product line development of avionics software. In *Software Product Lines* (pp. 353-369). Springer, Boston, MA.
- [32] Nagamine, M., Nakajima, T., & Kuno, N. (2016, September). A case study of applying software product line engineering to the air conditioner domain. In *Proceedings of the 20th International Systems and Software Product Line Conference* (pp. 220-226). ACM.
- [33] Hubaux, A., Classen, A., & Heymans, P. (2009, August). Formal modelling of feature configuration workflows. In *Proceedings of the 13th International Software Product Line Conference* (pp. 221-230). Carnegie Mellon University.
- [34] Czarnecki, K., Helsen, S., & Eisenecker, U. (2004, August). Staged configuration using feature models. In *International Conference on Software Product Lines* (pp. 266-283). Springer, Berlin, Heidelberg.
- [35] Salinesi, C., Mazo, R., Djebbi, O., Diaz, D., & Lora-Michiels, A. (2011, May). Constraints: The core of product line engineering. In *2011 FIFTH INTERNATIONAL CONFERENCE ON RESEARCH CHALLENGES IN INFORMATION SCIENCE* (pp. 1-10). IEEE.
- [36] Van Der Aalst, W. (2011). *Process mining: discovery, conformance and enhancement of business processes* (Vol. 2). Heidelberg: Springer.
- [37] Von Alan, R. H., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 28(1), 75-105.
- [38] Vaishnavi, V., & Kuechler, W. (2009). *Design research in information systems*, Association for Information Systems.
- [39] Mazo, R., Salinesi, C., Diaz, D., Djebbi, O., & Lora-Michiels, A. (2012). Constraints: The heart of domain and application engineering in the product lines engineering strategy. *International Journal of Information System Modeling and Design (IJISMD)*, 3(2), 33-68.
- [40] Salinesi, C., & Mazo, R. (2012). Defects in Product Line Models and how to identify them. *Software Product Line-Advanced Topic*, 97-122.
- [41] Salinesi, C., Mazo, R., & Diaz, D. (2010, May). Criteria for the verification of feature models. In *INFORSID 2010*.
- [42] Rincón, L., Giraldo, G., Mazo, R., Salinesi, C., & Diaz, D. (2015). Method to identify corrections of defects on product line models. *Electronic notes in theoretical computer science*, 314, 61-81.
- [43] Mazo, R., Salinesi, C., & Diaz, D. (2011). Abstract Constraints: A General Framework for Solver-Independent Reasoning on Product Line Models.
- [44] Khodabandelou, G., Hug, C., & Salinesi, C. (2014, May). A novel approach to process mining: Intentional process models discovery. In *2014 IEEE Eighth International Conference on Research Challenges in Information Science (RCIS)* (pp. 1-12). IEEE.
- [45] Epure, E. V., Ingvaldsen, J. E., Deneckere, R., and Salinesi, C. (2016, June). Process mining for recommender strategies support in news media. In *Research Challenges in Information Science (RCIS), 2016 IEEE Tenth International Conference on* (pp. 1-12). IEEE.
- [46] Deneckère, R., Hug, C., Khodabandelou, G., and Salinesi, C. (2014). Intentional process mining: discovering and modeling the goals behind processes using supervised learning. *International Journal of Information System Modeling and Design*, 5(4), 22-47.
- [47] Günther, C. W. (2009). *Process Mining in Flexible Environment*. PhD thesis, Technische Universiteit Eindhoven.

- [48] Schonenberg, H., Weber, B., Van Dongen, B., and Van der Aalst, W. (2008, September). Supporting flexible processes through recommendations based on history. In *International Conference on Business Process Management* (pp. 51-66). Springer, Berlin, Heidelberg.
- [49] Parthasarathy, S., Zaki, M. J., Ogihara, M., and Dwarkadas, S. “Incremental and interactive sequence mining.” in *Proceedings of the Eighth International Conference on Information and Knowledge Management*, ser. CIKM '99. New York, NY, USA: ACM, 1999, pp. 251–258.
- [50] Zhou, W., Liu, H., and Cheng, H. (2010, June). Mining closed episodes from event sequences efficiently. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 310-318). Springer, Berlin, Heidelberg.
- [51] Mazo, R., Muñoz-Fernández, J., Rincón, L., Salinesi, C., Tamura, G., (2015) *VariaMos: an extensible tool for engineering (dynamic) product lines*. In the *XIX International Software Product Line Conference (SPLC)*, Nashville-USA, July 20-24, 2015.
- [52] Günther, C. W., and Rozinat, A. (2012). Disco: Discover Your Processes. *BPM (Demos)*, 940, 40-44.
- [53] Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, xiii-xxiii.
- [54] Claes J., Poels G. (2013) *Process Mining and the ProM Framework: An Exploratory Survey*. In: La Rosa M., Soffer P. (eds) *Business Process Management Workshops. BPM 2012. Lecture Notes in Business Information Processing*, vol 132. Springer, Berlin, Heidelberg.
- [55] Ailenei, I., Rozinat, A., Eckert, A., van der Aalst, W.M.P.: Definition and Validation of Process Mining Use Cases. In: Daniel, F., Barkaoui, K., Dustdar, S. (eds.) *BPM Workshops 2011, Part I. LNBIIP*, vol. 99, pp. 75–86. Springer, Heidelberg (2012)
- [56] Basili, V.R., Caldiera, C., and Rombach, H.D.(1994) *Goal/Question/Metric Paradigm*,” *Encyclopaedia of Software Engineering*, Vol. 1,J.J. Marciniak, ed., John Wiley and Sons, New York, pp. 528-532.
- [57] Xiang, Y., Zhou, Y., Zheng, Z., & Li, M. (2018). Configuring software product lines by combining many-objective optimization and SAT solvers. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 26(4), 14.
- [58] Günther, C. W., & Van Der Aalst, W. M. (2007, September). Fuzzy mining–adaptive process simplification based on multi-perspective metrics. In *International conference on business process management* (pp. 328-343). Springer, Berlin, Heidelberg.
- [59] Benavides, D., Segura, S., & Ruiz-Cortés, A. (2010). Automated analysis of feature models 20 years later: A literature review. *Information systems*, 35(6), 615-636.
- [60] Hadzic T., Andersen H. R. An introduction to solving interactive configuration problems. Technical Report TR-2004-49, The IT University of Copenhagen, 2004.
- [61] Jannack, D., Zanker, M., Felfering A., & Friedrich G. (2011) *Recommender Systems, an Introduction*. Cambridge University Press. New York, USA, 335 pages.
- [62] Batory D. (2005) *Feature Models, Grammars, and Propositional Formulas*. In: Obbink H., Pohl K. (eds) *Software Product Lines. SPLC 2005. Lecture Notes in Computer Science*, vol 3714. Springer, Berlin, Heidelberg.
- [63] Mooney, R. J., & Roy, L. (2000, June). Content-based book recommending using learning for text categorization. In *Proceedings of the fifth ACM conference on Digital libraries* (pp. 195-204). ACM.
- [64] Linden, G., Smith, B., & York, J. (2003). Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, (1), 76-80.
- [65] Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64, 1-18.
- [66] Okoli, C., & Schabram, K. (2010). A guide to conducting a systematic literature review of information systems research.
- [67] Mendonca, M., Branco, M., & Cowan, D. (2009, October). SPLIT: software product lines online tools. In *Proceedings of the 24th ACM SIGPLAN conference companion on Object oriented programming systems languages and applications* (pp. 761-762). ACM.