Empirical software product line engineering: A systematic literature review

Ana Eva Chacón-Luna [a,*], Antonio Manuel Gutiérrez [b], José A. Galindo [c], David Benavides [c]

[a] *Universidad Estatal de Milagro, Km. 1 1/2 vía km. 26, Milagro, Ecuador*
[b] *ISIS Papyrus, Papyrus Platz 1, 2345, Brunn am Gebirge, Austria*
[c] *Dept. Lenguajes y Sistemas Informáticos, University of Seville, Avda. Reina Mercedes s/n, 41012, Seville, Spain*

ARTICLE INFO

ABSTRACT

Context: The adoption of Software Product Line Engineering (SPLE) is usually only based on its theoretical benefits instead of empirical evidences. In fact, there is no work that synthesizes the empirical studies on SPLE. This makes it difficult for researchers to base their contributions on previous works validated with an empirical strategy.
Objective: The objective of this work is to discover and summarize the studies that have used empirical evidences in SPLE limited to those ones with the intervention of humans. This will allow evaluating the quality and to know the scope of these studies over time. Doing so, research opportunities can arise
Methods: A systematic literature review was conducted. The scope of the work focuses on those studies in which there is human intervention and were published between 2000 and 2018. We considered peer-reviewed papers from journals and top software engineering conferences.
Results: Out of a total of 1880 studies in the initial set, a total of 62 primary studies were selected after applying a series of inclusion and exclusion criteria. We found that, approximately 56% of the studies used the empirical case study strategy while the rest used experimental strategies. Around 86% of the case studies were performed in an industrial environment showing the penetration of SPLE in industry.
Conclusion: The interest of empirical studies has been growing since 2008. Around 95.16% of the studies address aspects related to domain engineering while application engineering received less attention. Most of the experiments and case study evaluated showed an acceptable level of quality. The first study found dates from 2005 and since then, the interest in the empirical SPLE has increased.

1. Introduction

Software product lines (SPLs) aim to support the development of a family of software products through the systematic reuse of shared assets [1]. SPLs have received significant attention from software engineers since the 1990s because of the benefits they propose, such as reuse in software development, expansion of the product catalog, time reduction to market, increased return on investment, reduction of development costs, improved software quality or customer satisfaction among others [1–3]. Nowadays, SPLs have been widely developed in the academic world and applied in the industry. However, the application of SPLE is usually only based on its theoretical benefits, that is, its application is not always supported by empirical evidences [4].

We can find studies in the literature that describe the benefits of empirical strategies application in software engineering [5]. Depending on the technology used (technique, method or tool) and the purpose of the evaluation, there are strategies that can be used and adapted

to the particular needs. Each empirical strategy has its limitations and benefits.

According to Basili et al. [6], *the weaknesses and benefits of technologies can be identified through empirical studies*. This can explain why software engineering and more concretely, SPLE, has begun to emphasize empirical research methods when improving the validity and generalization of research results [7–10].

To the best of our knowledge, there is no work that discovers and summarize the empirical studies on the application of SPLE. To solve this need we have performed a Systematic Literature Review (SLR) in empirical evidences in SPLE. Literature Reviews are a method to identify relevant research, methods and gaps in existing research.

Concretely we chose relevant papers from four research databases (i.e. ACM, IEEE, Science Direct and Scopus) delimiting our study to those empirical studies that are completed with human intervention. This review focuses on SPLE empirical evidence with

Fig. 1. Phase the process of SLR [11,13].

Table 1
Research question.

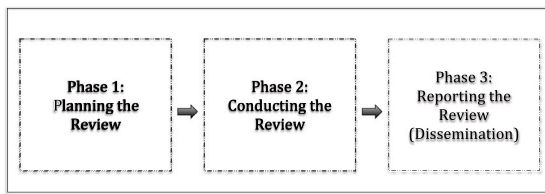| N°	Research questions
RQ1	Are there empirical studies on SPLE with human intervention?
RQ1.1.	What empirical strategies have been used?
RQ1.2.	What processes of the SPLE framework have been studied?
RQ1.3.	In what context is empirical research done?
RQ2	What is the quality of the empirical studies reported in SPLE?
RQ3	What is the scope of empirical research in SPLE?
RQ3.1.	In which journals or conferences are these types of articles published?
RQ3.2.	What is the temporal evolution in this type of studies?
RQ3.3.	What authors do research in this area?

human intervention between 2000 and 2018. In this period, we retrieved more 1800 peer-reviewed papers from journals and the top software engineering conferences in which authors published empirical research with human intervention. We filter these papers, such as those that are not short papers, to finally obtain 62 primary studies.

Our study is based on the definition of Kitchenham et al. [11], where it is indicated that the systematic reviews of the literature are carried out to *"identify, analyze and interpret all available evidence related to a specific research question"*.

The contributions of this paper are as follows:

1. We discover and summarize the current status of the studies which have used empirical evidence in SPLE with human intervention. We rely on the guidelines for carrying out systematic bibliographic reviews in software engineering presented by Kitchenham et al. [11], Kitchenham [12], Keele et al. [13].
2. We evaluate the maturity of the methods of the primary studies to highlight the limitations and strengths. The maturity of the experiments is analyzed according to Wholin guidelines [5] and, the one of case studies, with the ones of Runeson et al. [14] and Robson [15].
3. We show the evolution of these studies over time: the research topics addressed, where from and whom the authors are and the institutions that investigate in these topics.
4. We present the research gaps and highlights critical areas that require further study.

The rest of the paper is structured as follows. Section 2 describes the methodology of the systematic literature review. This methodology follows the guidelines presented by Kitchenham [12], evaluated in [13] and updated in [11]. Section 3 presents the results of the study. Section 4 presents related works and justifies our proposal and the results are discussed in Section 5 together with the conclusions.

2. SLR process definition

We have relied on the guidelines presented by Kitchenham [11–13] to perform systematic bibliographic reviews in software engineering. These guidelines are divided into three main phases: planning, conducting and reporting the review, as shown in Fig. 1.

In this section, we detail each step of the SLR.

2.1. Phase 1: Planning the review

The main objective of this phase is to develop a protocol to identify, evaluate and collate the evidence for the SLR. Fig. 2 details the steps that we followed and that are described in the following sections.

2.1.1. Identification of the need for a review

The main motivation of the execution of this systematic study is the need of synthesizing the empirical studies in SPLE to support the researchers for future contributions because, as reported in Section 4, no systematic review of the literature on empirical evidence was found, nor with the intervention of humans. The main objective of the review is to discover and summarize the studies that have used empirical evidences in SPLE limited to those ones with the intervention of humans, in order to evaluate the quality and to know the scope of these studies over time. Doing so, research opportunities can arise.

2.1.2. Specifying the research questions

The research questions direct the entire methodology of systematic review, i.e. the search process, data extraction and data analysis. Table 1 presents the research questions and sub-questions present in this study.

RQ1. Are there empirical studies on SPLE with human intervention?

We want to know the studies that have used empirical evidence in SPLE with human intervention. This question is answered with the following sub-questions.

- **RQ1.1. What empirical strategies have been used?** This research question seeks to know the empirical strategies that have supported the research conducted in SPLE. According to Robson [15] and Wohlin et al. [5], the empirical strategies are survey, case study, experiment or quasi-experiment.
- **RQ1.2. What processes of the SPLE framework have been studied?** This question seeks to know which processes of the SPLE framework given by Pohl et al. [2] have been researched by empirical studies. In this way, we determine the topics which have attracted the most attention of the researchers and show research gaps that can generate future works.
- **RQ1.3. In what context is empirical research done?** This question seeks to know what was the study environment (academic or industrial), the number of subjects who participated in the studies, the objective and the intention to use the empirical strategy.

RQ2. What is the quality of the empirical studies reported in SPLE?

Through this question we intend to know the quality of the empirical studies found. There could be different ways of measuring the quality of primary sources. However, we do it by measuring the maturity of the method used with well recognized proposals from Wohlin et al. [5], Runeson et al. [14] and Robson [15]. The maturity of the method represents how well a study was performed and reported according to the aforementioned guidelines. This allows us to perform an unbiased comparison between different contributions independently of the concept of quality that each reader might have. Note that in the paper we refer to the quality and maturity of the research method as synonyms.

RQ3. What is the scope of empirical research in SPLE?

We want to know the general data of the researches reported on empirical evidence in SPL and the evolution of these studies.

- **RQ3.1. In which journals or conferences are these types of articles published?** This question allows knowing the suitable publications for researchers who are interested in making future contributions in SPLE supported by empirical software engineering.
- **RQ3.2. What is the temporal evolution in this type of studies?** This question determines the temporary evolution of the research on SPLE. The objective is to help researchers to know about which topics and when the contributions were made.
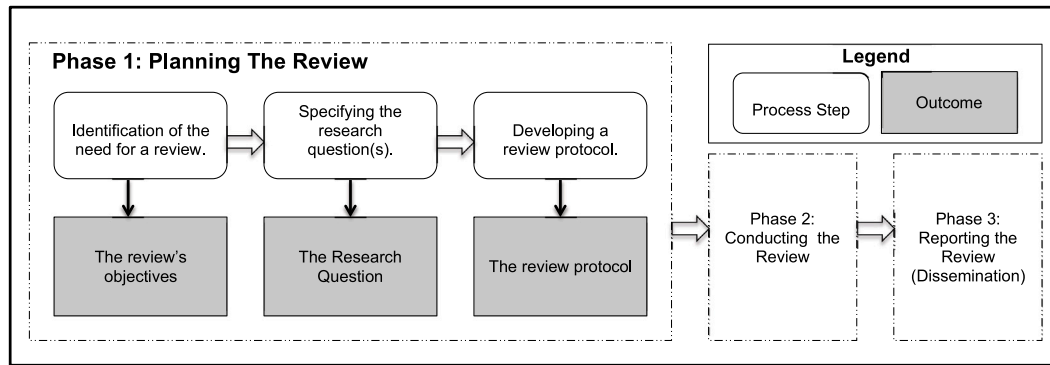
Fig. 2. Phase 1. Planing the review process [11,13].

Table 2
Keywords used to build the search strings.

Term	Keywords	Alternatives
A	Software Product Line	"Software Product Line" "Software Product Lines"
B	Software product family	"Software product family" "Software product families"
C	Empirical	"Empirical"
D	Survey	"Survey"
E	Experiment	"Experiment"
F	Case Study	"Case Study"
G	Experience Report	"Experience Report"

• **RQ3.3. What authors do research in this area?** This question allows knowing who are the authors and institutions which (supported by empirical studies) research on SPLs. In this way, it will be known which institutions have knowledge about the subject, leading to detect and encourage collaboration among researchers.

2.1.3. Developing a review protocol

The development of a review protocol allows specifying the steps that researchers follow during a review. Fig. 3 depicts the elements in this systematic review.

The first step is defining the set of research questions which this SLR intends to answer, as we have already described in Section 2.1.2. Next, we define the search strategy for primary studies. This step includes the definition of the search strings and the description of the digital libraries used in this work. We define the study selection criteria where the inclusion and exclusion criteria for the selection of the primary studies were specified. The next step is defining the study selection procedures. Here we describe how the selection criteria is applied, how each primary study is evaluated and how the disagreements between the researchers are resolved. The next step is the definition of checklists and procedures for study quality assessment. In this step, we justify the quality criteria applied to obtain the primary studies. Finally, we describe the data extraction procedure followed in this systematic literature review. Following, we detail each of the steps in the development of a review protocol.

Search for primary studies. Based on the research questions posed, we identify the keywords that should be used to conduct the primary studies search. Table 2 shows these keywords and the alternative synonyms.

The keywords were verified in relevant articles, such as [32,33]. Initial searches were also carried out with these keywords in the

databases. Alternative synonyms were incorporated using the Boolean 'OR'. In addition, the main terms were linked using the Boolean 'AND'.

This way, by combining the keywords A and B, we can find studies that focus on SPL, regardless how the authors name them in their work. The keywords C, D, E, F and G allow us to focus on works related to empirical strategies. In this case, we decided to include only the most general terms to precisely identify the primary studies. We also considered to include the name of the SPLE framework processes in the search string but after reviewing manually, all the relevant papers were already included by the proposed search string, so we decided not to include the processes to not increase the initial set with results that are not relevant and would hinders the replication of this study.

Following the strategy described, and after a series of test runs and revisions, we built the generic search strings as follows:

```
("Software product line" OR "Software product
lines"OR "Software product family"OR "Software
product families") AND ("Empirical" OR "Survey"
OR "Experiment" OR "Case study" OR "Experience
report")
```

We decided that our original search strategy would include two phases: automatic search of electronic databases and search by snowball reading. Whereas digital libraries have specific parameterizing for their interfaces, we realized that it was not possible to use a single search string for all bibliographic sources. Therefore, we decided to design and use different search strings, making the necessary efforts so these search strings were logically and semantically equivalent. As part of the planning, several test searches were conducted. The results of the test searches were discussed among the researchers to refine the search string until we were completely satisfied with the ability of the search string to obtain the primary studies. For more details, see Appendix A, where we present more details about the specific search strings used.

Study selection criteria. The following inclusion and exclusion criteria were defined, which would allow us to obtain the primary studies that we analyze in this paper. The criteria we decided to apply are listed in Table 3.

We considered in the study the articles published between 2000 and 2018 because in 2000 the software product line framework was established. We considered articles from journals as well as from the top 5 conferences with more results in the initial data. These 5 conferences are also high impact conferences in software engineering so the review process ensures enough mature and relevant papers. For the same reason we decided to eliminate those journal articles with nine pages or less and conference articles with five pages or less. After removing such papers we also discarded any other secondary study (which we have considered as related work) and those papers in which authors were the only participants in the study. This is done, not only to maintain the number of papers to review in a manageable amount but to keep on focus primary studies.
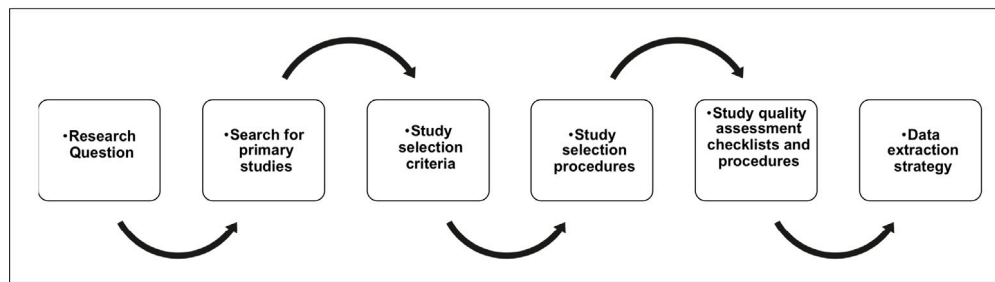
Fig. 3. Review protocol following Kitchenham [11].

Table 3
Inclusion and exclusion criteria.

Inclusion criteria
1. Papers published between 2000 and 2018
2. Peer reviewed articles
3. Papers related to empirical strategies with human intervention

Exclusion criteria
1. Peer reviewed articles of conferences, and workshops except from these conferences: •International Systems & Software Product Lines Conference (SPLC)   •International Symposium on Empirical Software Engineering and Measurement (ESEM)   •International Conference on Software Engineering (ICSE)   •International Symposium on Foundations of Software Engineering (ESEC/FSE)   •International Conference on Evaluation and Assessment in Software Engineering (ESEAM)
2. Papers from journals with nine pages or less and papers from conferences with five pages or less to exclude short papers.
3. Duplicate papers
4. Literature reviews, mapping studies, and tertiary reviews to focus only on primary studies.
5. Papers that do not have empirical evidence with intervention of humans in SPLE.
6. Papers where only the authors themselves are involved as participants

Study selection procedures. We rely on two bibliographic management tools: Mendeley[1] and Jabref.[2] Mendeley is used as a repository for the articles returned from the databases and Jabref for the application of the inclusion and exclusion criteria. We used Jabref tool because it provides a user-friendly interface for editing BibTeX files. We also relied on some automatic mechanisms to perform part of the bibliometric analysis such as the generation of figures. The application of inclusion and exclusion criteria is manually verified by all authors of this work. To apply the criteria "papers that do not have empirical evidence with intervention of humans", we agreed to read the title, abstract and conclusions in a first stage, and other key parts of the papers in the case that more information was needed. Following the guidelines from Kitchenham [11], the disagreements in the classification of the articles are resolved in a meeting to discuss them and reach a consensus.

Checklists and procedures for study quality assessment. Authors discussed the application of checklists to ensure the quality of the primary studies. We followed the guide by Tacconelli [16, Pag. 33] which indicates that: *"there is no single approach to assess quality. The importance of each aspect of quality will depend on the approach and nature of the review. The best approach will be determined by contextual, pragmatic and methodological considerations of the study"*. Furthermore, since one of our

research question was to assess the quality of the empirical studies (c.f. RQ2), we decided to not impose further restrictions to the papers in terms of quality.

Also, all the selected works were read thoroughly by the authors who approved them according to their experience for the inclusion as the primary sources in this list.

Data extraction strategy. The strategy agreed to extract data in this review consisted of reading the 62 primary studies and designing a series of forms that would help answer the research questions raised in the study. SPSS[3] and R[4] were used to classify and analyze the information obtained from primary studies. The specific forms used in this SLR are explained in Section 2.2.3.

2.2. Phase 2: Conducting the review

This phase executes the protocol defined in the previous phase. Fig. 4 details the actions performed during this phase.

2.2.1. Identification of research

The main activity in this step is to instantiate and apply the search string defined in for each chosen data source. Table A.19 of Appendix A shows each of the search strings used in the databases. The result column shows the number of studies obtained from each digital library. Our search in all the databases resulted in 2306 documents. All searches were based on the title, keywords and summary and the results were filtered between January 2000 and November 2018.

2.2.2. Selection of primary studies

The selection of our primary studies was the result of the application of the inclusion and exclusion criteria established in the planning stage. Fig. 5 shows each of the filters that were applied to obtain our primary studies.

Step 1: Conducting search. In this step we obtained 1880 papers. This is the union of the total of articles imported from the four databases to the bibliographic reference manager Mendeley. Note that in this step Mendeley already identified some duplicate items (426) which were automatically removed and, then, we exported the results to JabRef to continue with the application of the inclusion and exclusion criteria previously defined.

Step 2: Filtering studies. This step outputs the number of papers resulting after applying each of the six exclusion criteria previously defined in Table 3. The application of the first exclusion criteria generated discussions among the researchers. One the one side, following the factors recommended by Dybå [17], we considered excluding all the conference papers because their limited space can lead to incomplete papers. On the other side, the review process in top conferences ensures mature papers. Therefore, we finally agreed to consider the top 5 conferences with more papers in the results which also coincide with

---
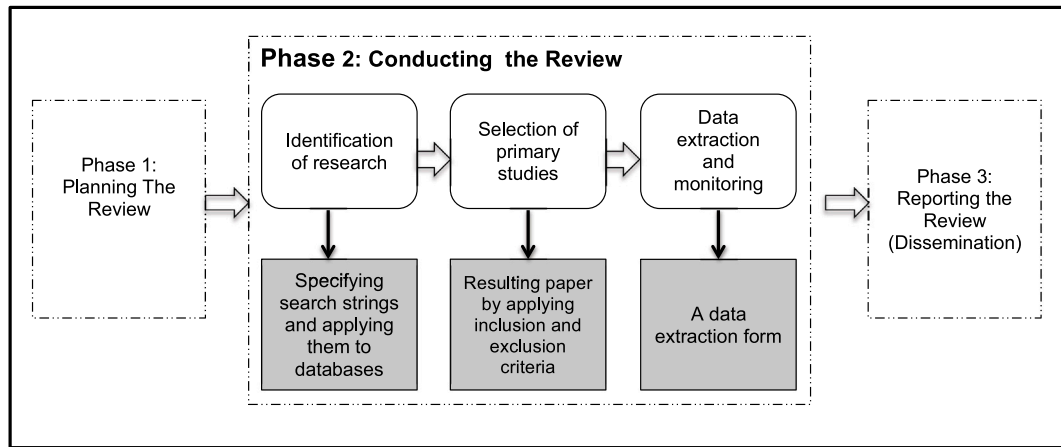
Fig. 4. Phase 2: Conducting the review [11].
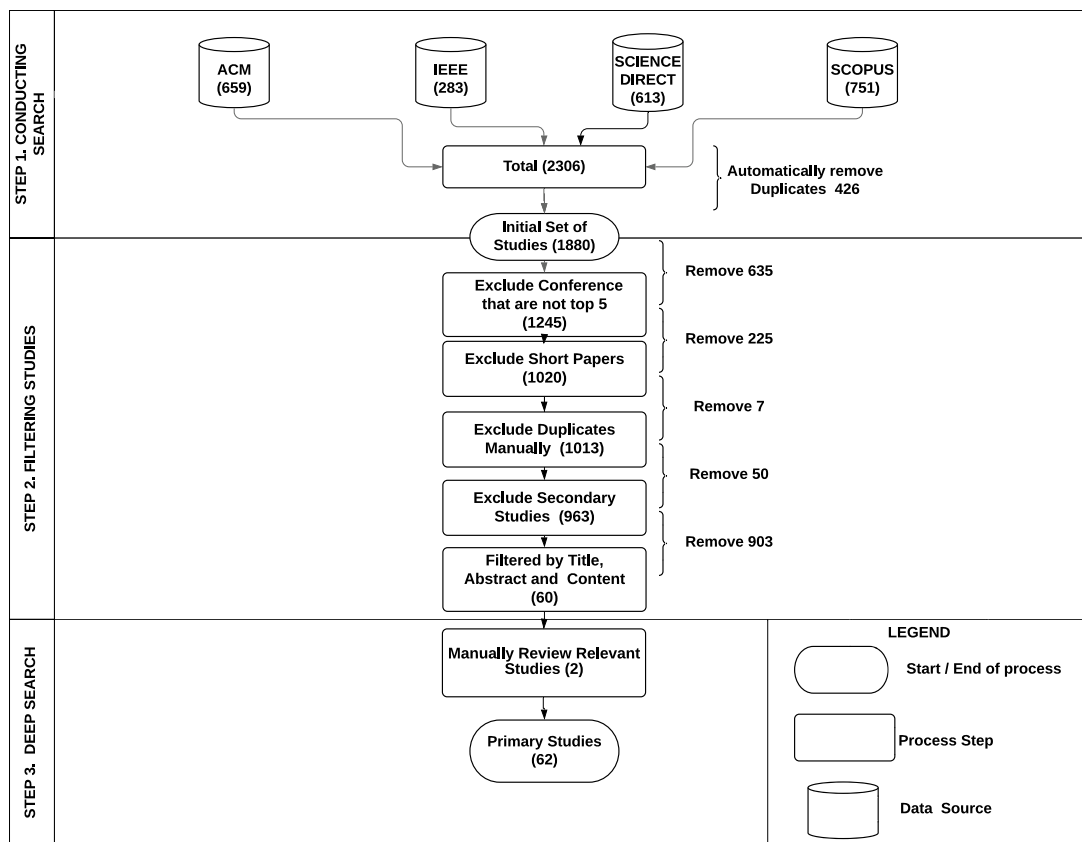
Fig. 5. Details of the selection phases during the selection of primary studies.

high impact conferences in the field of software engineering with product lines topics. We applied this exclusion criteria by eliminating 635 articles. In this stage we are left with a total of 1245 papers. Then we applied the second exclusion criteria, this is, we excluded all those articles from journals with less than ten pages and those articles from conferences with less than six pages. In this stage, a total of 225 articles were excluded, leaving us with 1020.

Then we applied the third exclusion criteria, which consisted of manually deleting all remaining duplicate items. It should be noted that, as it was already indicated in the first step, *conducting search*, some duplicates were already been removed automatically. Note that we also found an increment paper, that is a journal paper which extends the results of a conference papers which we also considered as a duplicate.

It was the case of [34] and [18] and we only added to our study the journal paper [34]. Therefore, in this stage, 7 articles were eliminated, leaving a total of 1013 articles.

The fourth exclusion criteria was applied to eliminate secondary studies and tertiary studies. A total of 50 articles were eliminated, leaving a total of 963 primary studies. The following exclusion criteria applied was based on the reading of titles, abstract and conclusions in order to determine whether the article met the topic of interest for this SLR. The researchers verified the fifth exclusion criteria, i.e. if it was an empirical study in SPLE with human intervention and if there were more participants in the study apart from the authors. When applying this criteria, 903 papers were excluded, resulting in 60 articles. None of these 60 articles included the authors as the only participants of the

study, as it was described in the sixth exclusion criteria, so this was the final list of articles for the analysis.

Step 3: Deep Search. To avoid losing any relevant work, we conducted a snowball reading search, in addition to the experience of the authors, to improve the final sample. We obtained two articles that met the inclusion criteria. This way we obtained a total of 62 primary studies from our SLR.

2.2.3. Data extraction and monitoring

The objective of this stage was to design data extraction forms for the collection of data from the primary studies. To avoid bias in the data extraction process, we tested our extraction forms with five primary studies randomly. After some tuning, we reached the conclusion that the final versions of the forms were adequate to extract the data for answering the research questions of our SLR.

In addition to extracting basic elements from articles such as the DOI, the title, the name of the authors, the year of publication or the paper, Table B.20 of Appendix B shows the general data properties and values that were used to extract and classify the information from the primary studies. Appendix B also contains Tables B.21 and B.22 that show the forms used for the extraction of data related to the quality of the method for the case studies and the experiments respectively (in order to answer the RQ2). The general process consisted on reading the papers and look for the corresponding data to extract. If the data was explicit and clear, it was included in the data extraction form; if it was clear but not included in a specific section/subsection, it was also processed and included; when the data was not clear although some implicit information was present, the data was not included.

2.3. Threads to the validity

The SLR that we present in this article aims to be as systematic as possible. However there are assumptions that we made that may affect its validity.

2.3.1. External validity

As mentioned above, we considered four scientific databases. This may have biased the process by missing some articles that are not in these databases. However, we minimize the impact of this threat by reviewing the references of each document included and verifying if we lack relevant research. The main threats to external validity are:

- Validity of the population. We reviewed a large number of documents to reduce the possibility of missing relevant jobs. When the reviewing of the title and abstract was inconclusive, the content of the articles was also read to make sure we did not leave out a relevant article.
- Ecological validity. This focuses on possible errors in the materials of the experiment and the tools used. We used an automated mechanism when possible, instead of relying on manual methods.

2.3.2. Internal validity

Another threat we faced was the bias of individual researchers when evaluating their assigned primary studies. To overcome this threat, we followed a predefined procedure and, specifically for the quality evaluation of the method used in the primary studies, we also were guided by predefined guides to develop a case study and an experiment.

3. Results

As a result of the analysis of the primary studies, we answer the research questions defined in Section 2.1.2.

Table 4 for case studies and Table 5 for experiments, show the aggregated data of the studies which are used to evaluate sub-questions RQ1.1, RQ1.2 and RQ1.3. The first nine columns indicate if they cover each of the nine processes defined in the SPLE Framework, while the

next five columns indicate the context of the work in order to answer RQ1.3.

The objective of the *Topic* column is to allow researchers to have information as a reference for future contributions. This value can help them to decide on which framework processes and on which specific topics new research should be reinforced or proposed. In general, there is a variety of topics studied such as Software Architecture, Feature modeling or Efficiency.

3.1. RQ1: Are there empirical studies on SPLE with human intervention?

3.1.1. RQ1.1: What empirical strategies have been used?

As shown in Tables 4 and 5, when analyzing the primary studies, we realized that approximately 56% of the studies used the empirical strategy of case studies (35 case studies). While the rest of the primary studies used an experimentation strategy (26 experiments and one quasi-experiment which is categorized as experiment [35]) were found.

We conjecture that a possible cause of the higher number of case studies is that industrial contexts are closer to the SPL community. This is, we can find multiple initiatives to bridge the gap between industry and the SPL research community such as an industrial track at SPLC conference which was promoted by the Software Engineering Institute from the Carnegie Mellon University.

3.1.2. RQ1.2: What processes of the SPLE framework have been studied?

We want to identify the SPLE framework processes that have been studied applying an empirical strategy. To answer this question, we use the definition of the framework by Pohl et al. [2]. The answer to this question give us insights about how many and which are the studies developed so far. In this way, researchers will be able to identify current trends. In addition, knowing these results will allow standing on previous works that have been validated using empirical strategies.

In Fig. 6 we can see that the process to which researchers have paid more attention is the Domain Requirement. Concretely, we have found 41 studies, subdivided into 28 case studies, twelve experiments and one quasi-experiment (66.13%). Other processes that also predominated were Domain Design with 25 studies (40.32%) and Product Management with 26 studies (41.94%).

Around 66.12% of the studies address topics related to Domain Requirements, while there are other areas, especially related to application engineering that have received less attention. We consider that this is due to the fact that, being the first process of the SPLE Framework and a core activity when managing software product lines, these activities are carried out either implicitly or explicitly in the different companies. For instance, most – if not all – the companies manage a product portfolio when running an SPLE strategy.

We observe that the framework processes that have had the least attention (with only seven articles) are the process of Domain Testing, Application Realization, Application Testing. The Domain Testing process is responsible for the validation and verification of the reusable components. The Application Realization process creates the application and the Application Testing process includes the activities necessary to validate and verify an application according to its specification. It is possible that less attention has been devoted to the aforementioned processes because they perform more complex tasks such as the creation and verification of applications. We consider that here there is an opportunity for researchers, since these processes, as they have not been taken care of, require more empirical research.

It should be noted that in some primary studies more than one SPL framework processes were evaluated. In such case they were accounted for all the activities that were covered, which means that the sum is not equal to the number of primary studies analyzed, 62.

3.1.3. RQ1.3: In what context is empirical research done?

We want to know which the study environment was (academic or industrial), the number of subjects that participated in the study, the objective and the intention to use the empirical strategy. Table 6 details these topics for the studied sources.

Table 4
Empirical studies on SPLE with case study.

#Paper	Product Management	Dom. Requir. Eng.	Domain Design	Domain Realization	Domain Testing	App. Requir. Eng.	Application Design	Application Realization	Application Testing	What is evaluated	Empirical objective	Study Environment	No. Participants	Type of sample	Topic
Ahmed [36]	✓	–	–	–	–	–	–	–	–	P	E	I	M	C	Organizational management, Org. behavior
Berger [37]	✓	✓	–	–	–	–	–	–	–	P	E	I	–	–	SPL, Variability Modeling, Empirical Soft. Engineering, Soft. Ecosystems
Classen [38]	–	✓	–	–	–	–	–	–	–	TO	S	I	–	–	Feature models, Code, Modeling, Language
Costa [39]	–	–	–	✓	–	–	–	–	–	TE	S	I	L	–	Ontology, Feature model
Da Silva [40]	✓	✓	–	–	–	–	–	–	–	P	E	I	L	C	Requir. engineering, Agile methods, SPL. scoping
Da Silva [41]	✓	✓	–	–	–	–	–	–	–	P	E	I	L	C	Agile, Multi-method approach
Deelstra [42]	✓	✓	✓	–	–	✓	✓	✓	–	P	E	I	–	–	Prod. derivation, Variability management
Deelstra [43]	✓	✓	–	–	–	–	–	–	–	P	S	I	–	–	Variability, Assessment, Evolution
Dey [44]	–	✓	–	–	–	–	–	–	–	P-TO	S	I	L	–	Requir. elicitation, Adaptive systems
De Souza [45]	✓	✓	✓	–	–	✓	✓	–	–	P	E	I	L	C	Prod. derivation
Díaz [46]	–	–	✓	–	–	–	–	–	–	P	S	I	L	–	Agile product-line engineering.
Díaz [47]	✓	✓	✓	–	–	–	–	–	–	TE	S	I	L	–	Traceability M., Prod.-line archit.,Variability
Echeverría [48]	–	–	–	–	✓	–	–	–	–	TE	E	I	L	–	SPL, Variability Modeling, Usability
Eklund [49]	–	–	✓	–	–	–	✓	–	–	P	E	I	L	–	Architecting, Process
Engström [50]	–	–	–	–	✓	–	–	–	✓	P	E	I	L	–	Soft. testing, Overlay, Redundancy, Efficiency
Eriksson [51]	–	✓	–	–	–	–	–	–	–	P	S	I	L	–	Feature model, Variability management
Ferreira [32]	✓	✓	✓	✓	✓	✓	✓	✓	✓	P	E	I	L	C	SPL adoption, Multi-method approach
Figueiredo [52]	✓	✓	–	–	–	–	–	–	–	TE	E	A	L	–	SPL, aspect-oriented programming, empirical evaluation.
Galster [53]	✓	✓	✓	–	–	–	–	–	–	P	E	I	L	–	Variability, Enterprise software systems
Ganesan [54]	✓	✓	✓	✓	✓	✓	✓	✓	✓	TE	E	I	L	–	SPL Applying Monte-Carlo Simulation
Hanssen [55]	✓	✓	✓	✓	✓	✓	✓	✓	✓	P	E	I	L	–	Soft. prod. development, Agile soft. development
Hanssen [56]	✓	✓	✓	✓	✓	✓	✓	✓	✓	P	E	I	L	–	Agile soft. development, Qualitative research
Kim [33]	–	✓	✓	–	–	–	–	–	–	P-TO	S	I	L	SR	Domain requir.,Domain archit.,Quality attribute
Koziolek [57]	✓	–	–	–	–	–	–	–	–	P	S	I	–	–	Domain analysis , Business case
Martínez [58]	✓	✓	–	–	–	✓	–	–	–	TE	E	A	L	–	SPL, Reverse-engineering, Extractive SPL Adoption
Myllärniemi [59]	✓	✓	✓	–	–	✓	✓	–	–	TE	S	I	L	C	SPL, variability, architecture
Oliveira [60]	✓	✓	–	–	–	–	–	–	–	P	S	A	L	C	Requir. Specification, Reuse
Pardo [61]	✓	✓	✓	✓	✓	✓	✓	✓	✓	P	E	I	L	–	Agile software development, Design thinking
Patzke [62]	✓	✓	✓	✓	–	–	–	–	–	P	S	I	–	–	SPL Code Evolution, Goal-Based SPL Measurement, Variability Code Smells
Rubin [63]	✓	✓	✓	✓	✓	–	–	–	–	P	E	I	–	–	Legacy SPL, cloned product variants.
Souza [64]	✓	✓	–	–	–	–	–	–	–	TE	E	I	L	C	Soft. quality control, Soft. inspection
Thurimella [65]	✓	✓	–	–	–	–	–	–	–	TE	S	A	H	SR	Variability management,variability modeling
Usman [66]	–	✓	–	–	–	–	–	–	–	P-TO	S	I	–	–	Mobile applications, Feature model
Wille [67]	✓	✓	–	–	–	–	–	–	–	TE	S	I	L	–	Variability mining, technical architecture, enterprise architecture
Zhang [68]	–	✓	–	–	–	✓	–	–	–	TE	S	A	L	–	Non-functional requir. framework, Feature M.

What is evaluated: TE = Technique; P = Process; TO = Tool.
Empirical objective: S = Proposal from the same paper; E = Evaluates an existing proposal; R = Replica of an experiment.
Study Environment: A = Academic; I = Industrial.
No. Participants: L = No. participants < 20; M = 20 ≤ No. participants < 100; H = No. participants ≥ 100; - = Not specified.
Type of sample: SR = Simple random sampling; ER = Stratified random sampling; C = Convenience sampling; - = Not specified.

Fig. 6. Strategy of the studies according to the processes.

Table 5
Empirical studies on SPLE with experiment.

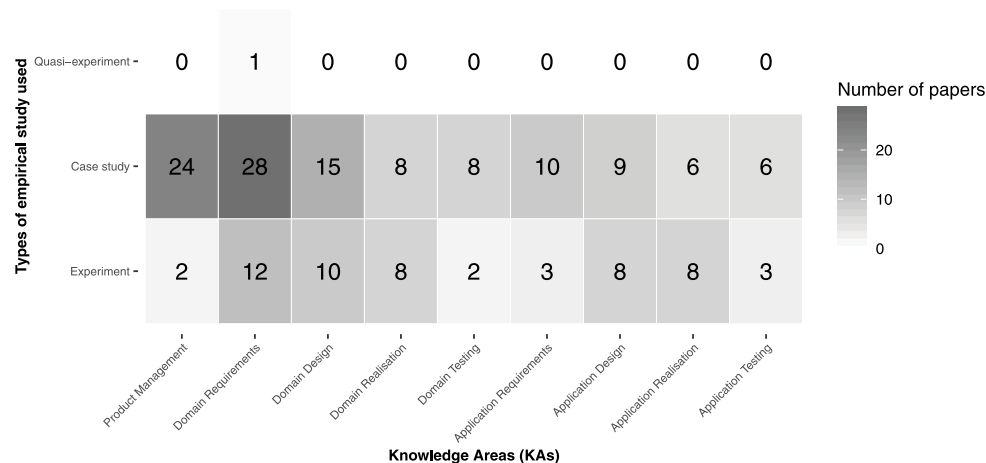| #Paper	Product Management	Dom. Requir. Eng.	Domain Design	Domain Realization	Domain Testing	App. Requir. Eng.	Application Design	Application Realization	Application Testing	What is evaluated	Empirical objective	Study Environment	No. Participants	Type of sample	Topic
Accioly [69]	–	–	–	–	✓	–	–	–	✓	TE	R	A	M	SR	Black-box Testing
Ahmed [70]	–	–	✓	–	–	–	✓	–	–	TE	E	I	M	C	Software archit.,Domain engineering
Asadi [71]	–	✓	–	–	–	–	–	–	–	TE	E	A	L	SR	FM Tools, Visualization and Interaction
Bagheri [72]	–	✓	–	–	–	✓	–	–	–	TE	S	A	L	–	Feature M., Quality attributes, Soft. prediction.
Bonifácio [73]	–	✓	–	–	–	–	–	–	–	TE	E	A	M	C	Requir. engineering, Software modularity
Cetina [74]	–	–	✓	–	–	–	✓	–	–	TE	E	A	–	–	Dynamic SPL,Runtime configuration
Denger [75]	–	–	–	–	–	–	–	✓	✓	TE	S	A	L	–	SPL, Reusable Components, Quality Assurance, Functional Testing
Dermeval [76]	–	✓	–	–	–	–	–	–	–	TE	E	A	L	SR	Feature modeling,Ontology
Feigenspan [34]	–	–	–	✓	–	–	–	–	–	TE	S	A	M	C	Software visualization, Program comprehension
Gonzalez-Huerta [77]	–	–	✓	–	–	–	✓	–	–	TE	S	A	M	SR	Software Archit., Quality attributes
Guana [78]	–	✓	✓	✓	–	–	–	–	–	TE	S	I	–	–	Soft. architecture, Variability management
Kumar [35]	–	✓	–	–	–	–	–	–	–	TE	E	A	H	ER	Variability, Requir. engineering, Mixed-methods
Liu [79]	✓	–	–	–	–	–	–	–	–	TE	S	A	M	SR	Domain analysis, Feature extraction
Martínez [80]	–	–	–	✓	–	–	✓	✓	–	TO	S	A	L	C	SPL Engineering, Reverse Engineering,Mining existing assets
Michalik [81]	–	–	–	–	–	–	–	✓	–	TE	S	I	L	C	SPL, on-line updates, experimentation
Pereira [82]	–	–	✓	✓	–	–	✓	✓	–	TE	S	A	M	C	SPL, Runtime Decision-Making, Self-Configuration,Recommender Systems.
Reinhartz-Berger [83]	–	✓	✓	–	–	–	–	–	–	TE	S	A	M	ER	Variability analysis, Model Comprehension, Empirical Study
Reinhartz-Berger [84]	–	✓	✓	–	–	–	–	–	–	TE	E	A	H	–	Variability management, Domain models.
Reinhartz-Berger [85]	–	✓	–	–	–	–	–	–	–	TE	E	A	M	SR	Variability modeling, Feature modeling
Saeed [86]	–	✓	–	–	–	–	–	–	–	TE	E	A	M	SR	Feature models, Visual syntax evaluation
Santos [87]	–	–	–	✓	–	–	–	✓	–	TE	S	I	L	C	SPL Engineering, Web Systems Domain, Feature Composition, FeatureIDE
Schulze [88]	–	✓	–	–	–	–	–	–	–	TO	S	A	L	–	Dynamic SPL, Variability modeling
Sinnema [89]	✓	✓	✓	✓	✓	✓	✓	✓	✓	P	E	I	L	–	Soft. Variability Management, Ind. validation
Stein [90]	–	✓	✓	–	–	–	–	–	–	P	S	A	M	C	Feature Model Configuration, Preferences
Thüm [91]	–	–	–	✓	–	✓	✓	–	–	P	S	A	M	C	SPL, preprocessor variability, evolution, maintenance.
Vasilevskiy [92]	–	–	–	–	–	–	–	✓	–	P	S	A	L	C	Realization, Product derivation, bvr; product line, model.variation point
Wang [93]	–	–	✓	✓	–	–	✓	✓	–	TE	S	A	M	SR	User modeling, Product line archit.

What is evaluated: TE = Technique; P = Process; TO = Tool.
Empirical objective: S = Proposal from the same paper; E = Evaluates an existing proposal; R = Replica of an experiment.
Study Environment: A = Academic; I = Industrial.
No. Participants: L = No. participants < 20; M = 20 ≤ No. participants < 100; H = No. participants ≥ 100; - = Not specified.
Type of sample: SR = Simple random sampling; ER = Stratified random sampling; C = Convenience sampling; - = Not specified.
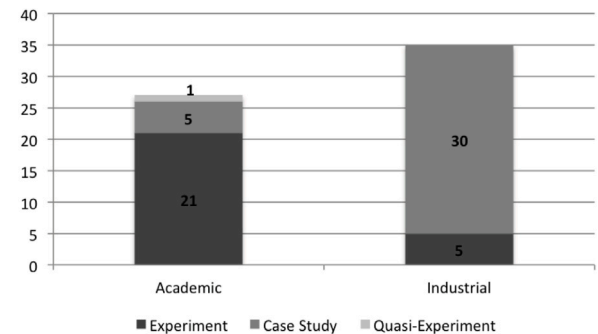
Table 6
Element of context.

Elements of context	Values
Study environment	-Academic -Industry
Number of participants	No.
What is evaluated	-Technique -Method -Tool -Method and Tool
Intention of the applied method	-Evaluate a proposal of the same article -Evaluate an existing proposal -It is a replica

Fig. 7. Environment of study.

Study environment. This aspect of the context aims to determine if the environment where the studies have been performed is academic, industrial or both.

In Fig. 7, we group the data in Tables 4 and 5 to illustrate that most case studies were conducted in an industrial context while most experiments were conducted in an academic context. It can be deduced that the SPLE increasingly gains ground in the industrial area. While 30 case studies were carried out in an industrial setting, only five were carried out in an academic environment. However, when analyzing the experiments, we observed that only five experiments were carried out in an industrial context, while 22 were carried out in an academic context. This behavior can respond to several factors. One of them is the design type of the empirical strategy used. According to Wholin [5, Pag.12], the experiments are more rigid at the time of execution with respect to the flexibility than a case study. In addition, the experiments demand more time and resources, both human and economic. It is likely that these factors are the reason why industries or the researchers themselves have desisted to conduct more experiments in an industrial context.

Number of participants. This element of the context provides the number of people who participated in the studies analyzed. For the present study we recognize as participants those individuals that were observed or were part of the primary study that we are analyzing; either a case study, experiment or quasi-experiment

Table 7
Number of participants.

Range	Participants	Papers	Percentage
L	≤20	35	56,45%
M	>20 and ≤100	14	22,58%
H	>100	3	4,84%
–	Not specified	10	16,13%
	TOTAL	62	100%

Table 8
Sampling type.

Sampling	Experiment	Case Study	Quasi-Experiment	Frequency	Percentage
Simple random	8	2	0	10	16,12%
Stratified random	1	0	1	2	3,23%
Convenience	11	8	0	19	30,65%
Not specified	6	25	0	31	50%
TOTAL	26	35	1	62	100%

Table 9
What is evaluated.

Option	Experiment	Case Studies	Quasi-Experiment	Total
Technique	20	11	1	32
Method	4	20	0	24
Tools	2	1	0	3
Process and Tools	0	3	0	3
TOTAL	26	35	1	62

Table 10
Empirical objective.

Option	Experiment	Case Studies	Quasi-Experiment	Total
Evaluates a proposal from the same paper	16	16	0	32
Evaluates an existing proposal	9	19	1	29
Is a replica of an experiment	1	0	0	1
TOTAL	17	25	1	62

As observed in Table 7, we defined three ranges of numbers of participants. In 10 studies of the total sample, the number of participants was not specified. We can also observe that 35 primary studies have worked with a small sample for the execution of the empirical strategy. Only three studies, that represent 4.8% of the total, used a sample superior or equal to 100 participants. We believe that this sample selection behavior was motivated by working with a representative sample, so they used subjective criteria based on the research they wanted to perform. This criteria was proposed to justify the number of participants in their research.

We also extracted information about the type of sample used in the investigations. Table 8 shows that 50% of the studies, i.e. 31 studies, do not indicate how they obtained the sample. We believe that this aspect must be improved. It should be adequately justified as they chose the participants to generalize the answers of the study. Most of experiments and case studies chose the sample for convenience, although 50% of the works did not indicate the sampling.

What is evaluated by applying empirical strategies?. According to Basili [6,19], the development of new techniques, methods and tools is one of the main concerns of the research in software engineering. The weaknesses and benefits of such technologies can be identified by conducting empirical studies.

Based on the contributions found in the primary studies, we confirm this statement, that is, the empirical strategies applied in SPLE aim to evaluate the proposals of new techniques, methods or tools.

Table 9 shows that most of the experiments were responsible for evaluating techniques, obtaining 20 primary studies that perform this task. The case studies mostly served as support to evaluate methods (20 primary studies). It should be noted that in three primary studies the case studies were responsible for the evaluation of two elements presented in the article, i.e. through the case study evaluated methods and tools.

We observe that researchers mostly preferred to evaluate their techniques using experiments. We consider that this behavior responds to the fact that the experiment generates exact quantitative data and allows them to explain or prove statistically if the used techniques solve a situation. While researchers preferred to execute case studies when evaluating methods, since the qualitative data that is collected allows identifying key factors that have an impact on the evaluated method.

What was the intention of the applied method?. We use the three possible objectives of the applied method (found in the primary sources), which are shown in the rows in Table 10, while the empirical strategies used are shown in the columns. In this way we have that only one of the reported experiments was part of a replica while most of the case studies (19) and nine experiments evaluated an existing proposal. Most of them aimed to reaffirm important findings in previously presented works. Another part of the case studies (16) and experiments (16) aimed to evaluate a proposal given in the same article. That is, these studies proposed a new technique, method or tool respectively to address some aspect of the SPLE and through the empirical strategy they evaluated how feasible or efficient their proposal was.

Key findings

- The empirical strategy mostly used in SPLE subjects is the case study with approximately 56% while the rest of the primary studies used experimentation.
- Around 95.16% of the studies address issues related to domain engineering, while application engineering received less attention.
- 86% of the case studies were performed in an industrial environment.
- 81% of the experiments were conducted in an academic environment.
- 50% of the primary studies do not specify the type of sample that was used to determine the study participants.

3.2. RQ.2.What is the quality of the empirical studies reported in SPLE?

There could be different ways of measuring the quality of primary sources. We also use known guides that should be followed in empirical software engineering. We have relied on the Wholin guidelines [5] to evaluate the maturity of the method of the experiments and on Runeson et al. [14] and Robson [15] to evaluate the case studies. Note that, we have considered the only quasi-experiment as an experiment to evaluate its maturity because it was developed in an academic environment.

Quality of the experiment method. The results are depicted in Table 11 with the 15 criteria that the experiments must meet to have quality together with the used method. Each criteria can receive a boolean value. Table 11 can be analyzed in two ways. On the one side, we can review how many works meet each criteria over a total of 27 experiments.

We observe that the most neglected criteria among the analyzed experiments (used in just 25.93% of the works) is the *Data Reduction*. This criterion seeks to eliminate those outliers that can invalidate a completely valid relationship.This might happen also, because most of the found studies, have a low number of participants. In such studies is difficult to determine if some observations are outliers and, moreover, discarding some of them would imply to discard a large percentage of the used dataset.

We also note that the *Variables Selection* criteria has been applied in 70.37% of the works. This allows us to choose dependent and independent variables. This criterion should receive more attention because it includes the choice of measurement scales, the range for the variables and the specific levels at which the tests will be carried out. The importance of this criteria affects the validity of the different studies, and therefore, authors may have taken it for granted, doing it informally and not reporting it properly. Other criteria that were used in less than 80% of the works were *Hypothesis Evaluation*, *Validity Evaluation* and *Hypothesis Testing*. They are intrinsic elements of the experiment although they have not been reported adequately in the primary studies.

On the other side, Table 11 also lists in the last column how many criteria meet each work. We see that the experiments that obtained the least rating (in terms of number of matched criteria) were [80] and [92], that met only three criteria and [78] and [74], that met seven and eight criteria, respectively. Note that the maximum possible rating is the number of criteria, 15. When reviewing the details about the publication of these experiments, we realized that the experiment [80] and [92] were published in conferences so that could explain the experimentation was not enough mature and the experiments [74] and [78] were published in 2013 and 2012, respectively, i.e. more than 5 years ago. Since the use of experimentation in SPLE was not as mature in those years, it is probable that they had some deficiencies in the method used for the execution of the experiment.

Quality of the case study method. The detailed results for the case studies are shown in Table 12 where the columns show the 10 criteria that the case studies must meet to have quality. As for the criteria of the experiments, here it is possible to receive a boolean value with the same connotation as in Table 11.

We observe that one of the criterion applied by the lowest number of studies, 8 (22.86% of the works), are *Propositions and Hypothesis*. This is, there are several studies in which, neither design nor execution of their case study, had verifiable empirical statements. Another criterion that was only applied by eight studies is *Methods of Data Analysis*. Since the case studies analyze qualitative data, the steps taken to analyze these data should be shown. That is, the codes used by the researcher must be shown for data analysis.

Also, the *Research Question* criterion has been applied in 20 studies (57.14% of the works). It draws our attention, because we consider that the realization of a case study requires not only expressing the intention of the research (in terms of objective) but also the research questions. Another criterion that researchers have overlooked is *Threats to validity*, included in 26 studies (74.29% of the works). We consider that in a case study the application of this criteria is very important because the case studies, having a flexible design, must minimize the effects of the threat in order to bring greater seriousness and reliability to the study.

The *Selection of Data* criterion was applied by 31 studies (88.57% of the works). This criterion seeks to specify by whom and when the study will be conducted. The objective is to ensure sufficient coverage of the data sources, this will support the triangulation of the data and increase the reliability and validity of the study findings.

The case studies are arranged in Table 12 in a descending order according to the number of criteria that they met. The case studies that obtained the least qualification were [61], [52] and [63] which met only four criteria. The case studies [38], [43], [54], [58] and [66] met five criteria. We consider that the case study design has not been enough specific, or at least it has not been reported in the primary studies and has left aside important aspects.

As in the evaluation of the experiment, Table 12 shows the most rated case studies according to matched criteria so it could serve as a guide for the use of an adequate method. We recommend to pay more attention to the elements that other researchers overlooked, with the aim of correcting these gaps in their contributions.

We evaluated the quality of the presented summary as a complement to the evaluation of case studies. We are inspired by Robson's work [15] where he specifies the characteristics that a case study report should have, shown in Table B.23 of the Appendix B.

After analyzing the case studies and verifying if they met the characteristics outlined by Robson for the presentation of case study reports, we obtained the results shown in Table 13. We can observe that the studies that obtained the lowest score are similar to the qualification obtained in Table 12 where the quality of the employed method was treated. We recommend that, for the evaluation obtained in the presentation of case study reports, the data should be provided in a focused form, so the reader can have confidence in the conclusions (C4) and these conclusions should be better articulated according to the context they affect (C5).

Finally, we highlight that only 59.26% of the experiments (14 or more criteria fulfilled) and 54.29% of the case studies (8 or more criteria fulfilled) obtained an adequate rating when evaluated. We conjecture that this low rating can be caused because of the disconnection between the SPL community and the empirical software engineering [20]. Fortunately, there are recent efforts that try to bridge that gap. For example, SPLC'19 and SPLC'10 special issues are being edited at Empirical Software Engineering journal. Also, the workshop on empirical SPLE that is being run in the last editions of the conference [21]. We expect that with such actions, the studies to be reported take these guidelines into account.

- 18.52% of the experiments met all the criteria recommended by Wholin [5]
- 37.14% of the case studies met nine of the 10 criteria recommended by Runeson [14]
- 49% of the studies show weakness in providing the history of the inquiry (C3)

3.3. RQ.3. What is the scope of empirical research in SPLE?

3.3.1. RQ3.1. In which journals or conferences are these types of articles published?

This question allows us to know and point out the most suitable journals or conferences for researchers who are interested in making future contributions in SPLE supported by empirical software engineering. Table 14 shows the list of journals and conferences where the papers have been published and the empirical strategies used (with the acronym EXP for experiment, CS for case study and QUASI for quasiexperiment). We can observe that the preferred sources by the authors when making a contribution have been the International Systems & Software Product Line Conference and the journals Journal of Systems and Software, and Information and Software Technology. It is worth mentioning that those conferences do not specifically target empirical research and are more focused on software engineering. Oppositely the venues that have their main focus in empirical software engineering do not publish many product line related contributions.

Table 11
Experiment summary.

#Paper	Goal Definition	Context Selection	Hypothesis Formulation	Variables Selection	Selection of Subjects	Experiment Design	Instrumentation	Validity Evaluation	Preparation	Execution	Data Validation	Descriptive Statistics	Data Reduction	Hypothesis Testing	Presentation and Package	TOTAL
Accioly [69]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	15
Gonzalez-Huerta [77]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	15
Kumar [35]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	15
Reinhartz-Berger [84]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	15
Reinhartz-Berger [85]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	15
Asadi [71]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	14
Dermeval [76]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	14
Feigenspan [34]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	14
Michalik B. [81]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	14
Ahmed [70]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	14
Bagheri [72]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	14
Thüm T. [91]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	14
Wang [93]	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	14
Denger C. [75]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	14
Saeed [86]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	14
Santos A. [87]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	14
Bonifácio [73]	✓	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	13
Liu [79]	✓	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	✓	-	✓	✓	13
Reinhartz-Berger [83]	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	13
Schulze [88]	✓	✓	✓	-	✓	✓	-	✓	✓	✓	✓	✓	-	✓	✓	12
Pereira J. [82]	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	11
Stein J. [90]	✓	✓	-	-	✓	✓	✓	-	✓	✓	✓	✓	-	-	✓	10
Sinnema [89]	✓	✓	✓	-	✓	-	-	-	✓	✓	✓	-	-	✓	✓	9
Cetina [74]	✓	✓	✓	-	✓	-	✓	-	✓	✓	-	-	-	-	✓	8
Guana [78]	✓	✓	-	-	✓	-	-	-	✓	✓	✓	-	-	-	✓	7
Martínez J. [80]	✓	✓	-	-	-	-	✓	-	-	-	-	-	-	-	-	3
Vasilevskiy A. [92]	✓	✓	-	-	-	-	-	✓	-	-	-	-	-	-	-	3
TOTAL	27	27	21	19	25	22	22	21	25	25	24	22	7	20	24	331

3.3.2. RQ3.2. What is the temporary evolution in this type of studies?

This data allows researchers to know from which year more empirical contributions have been made in SPLE.

Fig. 8 shows that the number of empirical studies has grown since 2004. We consider that this evolution responds to the need of evaluation and validation of the recent research. In addition, as mentioned by Wholin et al. [5], more empirical strategies in software engineering are now widely recognized than ten years ago. In Fig. 9 we observe that in 2013 the number of case studies increased and had no much variation since then. Note that after filtering, no article from 2010 was included in the primary sources.

Also, to have more details about this evolution, we present these studies classified by empirical strategy and the year of publication in Table 15 and Fig. 9. In Fig. 9, we observe that both experiments and case studies have been increased the number of publications from 2010 to 2013, have maintained the publications between 2013 and 2017 and have decreased in 2017 and 2018.

3.3.3. RQ3.3. What authors do research in this area?

This question allows knowing the authors and institutions that investigate SPLE supported by empirical studies. In this way, it will be known which institutions have knowledge of the subject, leading to detect and encourage collaboration among researchers.

Table 16 lists the number of empirical studies per author (with more than one paper authored) together with a heat map for the number of times that each process of the framework is addressed (using their acronym) by an author. Again, the processes with more studies are Product Management, Domain Requirement and Domain Design. This heat map also highlights how many times an author has used each experimental strategy.

Table 12
Case study summary.

#Paper	Rationale	Objectives	Cases and Units of Analysis	Theoretical Frame of Reference	Research Questions	Propositions and Hypothesis	Methods of Data Collection	Methods of Data Analysis	Selection of Data	Threats to Validity	Total
Ahmed F. *et al.* [36]	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	9
Da Silva IF. *et al.* [40]	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	9
Dey S. *et al.* [44]	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	9
De Souza LO. *et al.* [45]	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	9
Díaz J. *et al.* [46]	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	9
Díaz J *et al.* [47]	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	9
Engström E. *et al.* [50]	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	9
Eriksson M. *et al.* [51]	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	9
Galster M. *et al.* [53]	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	9
Koziolek H. *et al.* [57]	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	9
Oliveira RPD. *et al.* [60]	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	9
Costa GCB. *et al.* [39]	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	9
Ferreira J. *et al.* [32]	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	9
Souza IS. *et al.* [64]	✓	✓	✓	✓	✓	-	✓	-	✓	✓	8
Da Silva IF. *et al.* [41]	✓	✓	-	✓	✓	-	✓	✓	✓	✓	8
Berger T. *et al.* [37]	✓	✓	✓	✓	✓	✓	✓	-	✓	-	8
Echeverría J. *et al.* [48]	✓	✓	✓	✓	✓	-	✓	-	✓	✓	8
Hanssen GK. *et al.* [55]	✓	✓	✓	✓	-	-	✓	✓	✓	✓	8
Myllärniemi V. *et al.* [59]	✓	✓	✓	✓	✓	-	✓	-	✓	✓	8
Kim J. *et al.* [33]	✓	✓	✓	✓	-	-	✓	-	✓	✓	7
Hanssen GK. *et al.* [56]	✓	✓	✓	✓	-	-	✓	-	✓	✓	7
Thurimella A. *et al.* [65]	✓	✓	✓	✓	-	-	✓	-	✓	✓	7
Eklund U. *et al.* [49]	-	✓	✓	✓	-	-	✓	-	✓	✓	6
Deelstra S. *et al.* [42]	✓	✓	✓	✓	-	-	✓	-	✓	-	6
Zhang G. *et al.* [68]	✓	✓	✓	✓	-	-	-	-	✓	✓	6
Patzke T. *et al.* [62]	✓	-	✓	✓	-	-	✓	-	✓	✓	6
Wille D. *et al.* [67]	✓	✓	-	✓	✓	-	-	-	✓	✓	6
Classen A. *et al.* [38]	✓	✓	-	✓	✓	-	✓	-	-	-	5
Deelstra S. *et al.* [43]	✓	✓	✓	✓	-	-	✓	-	-	-	5
Ganeson D. *et al.* [54]	✓	✓	✓	✓	-	-	-	-	✓	-	5
Martinez J. *et al.* [58]	✓	✓	✓	✓	-	-	✓	-	-	-	5
Usman M. *et al.* [66]	✓	✓	✓	-	-	-	-	-	✓	✓	5
Pardo A. *et al.* [61]	-	✓	✓	✓	-	-	✓	-	-	-	4
Figueiredo E. *et al.* [52]	✓	✓	-	✓	-	-	-	-	✓	-	4
Rubin J. *et al.* [63]	✓	-	✓	✓	-	-	-	-	✓	-	4
TOTAL	33	33	31	34	20	8	29	8	31	26	253

Table 17 shows the names of the authors together with their affiliations. It is worth mentioning that the authors had low publication continuity on the same topics. Therefore, although there is a growth in this type of contributions, few authors lead in this area of research. This might be caused due the time span required to report a case study or perform experiments, which might take years according to our experience and therefore the publication throughput is lower in general. The authors from Brazil, Germany and Spain are the ones with more publications. Fig. 10 provides more details. It is interesting to note that SPLE empirical investigations are performed mostly in Europe although Brazil has a big amount of contributors too.

- The journals preferred by the authors when making a contribution with empirical components in SPLE have been the Journal of System and Software, Information and Software Technology and the conference preferred has been the International Software Product Line Conference
- The evolution of the studies has been growing, i.e. empirical strategies in software engineering are more recognized today than ten years ago.
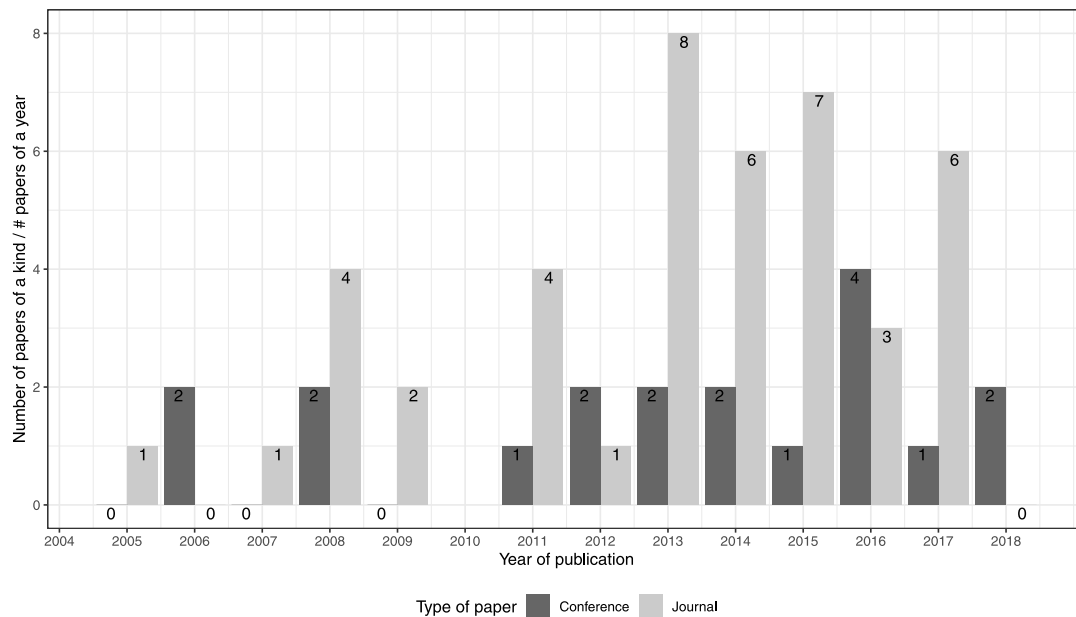
Fig. 8. Publication chronology of the papers included in review per year.

Table 13
Characteristics that a case study report should have [15].

#Paper	C1	C2	C3	C4	C5	Total
Ahmed F. et al. [36]	1	1	1	1	1	5
Da Silva IF. et al. [40]	1	1	1	1	1	5
Díaz J. et al. [46]	1	1	1	1	1	5
Echeverría J. et al. [48]	1	1	1	1	1	5
Eriksson M. et al. [51]	1	1	1	1	1	5
Galster M. et al. [53]	1	1	1	1	1	5
Ganeson D. et al. [54]	1	1	1	1	1	5
Koziolek H. et al. [57]	1	1	1	1	1	5
Martinez J. et al. [58]	1	1	1	1	1	5
Millärniemi V. et al. [59]	1	1	1	1	1	5
Oliveira RPD. et al. [60]	1	1	1	1	1	5
Rubin J. et al. [63]	1	1	1	1	1	5
Wille D. et al. [67]	1	1	1	1	1	5
Dey S. et al. [44]	1	1	1	1	0	4
Díaz J et al. [47]	1	1	1	1	0	4
Engström E. et al. [50]	1	1	0	1	1	4
Ferreira J. et al. [32]	1	1	0	1	1	4
Figueiredo E. et al. [52]	1	1	0	1	1	4
Kim J. et al. [33]	1	1	0	1	1	4
Berger T. et al. [37]	1	1	0	0	0	2
Classen A. et al. [38]	1	1	0	0	1	3
Costa GCB. et al. [39]	1	1	1	0	0	3
Da Silva IF et al. [41]	1	1	0	0	1	3
De Souza LO. et al. [45]	1	1	0	0	1	3
Deelstra S. et al. [42]	1	1	0	0	1	3
Hanssen GK. et al. [56]	1	1	0	0	1	3
Eklund U. et al. [49]	1	1	0	0	1	3
Pardo A. et al. [61]	1	1	0	0	1	3
Patzke T. et al. [62]	1	1	0	1	0	3
Thurimella A. et al. [65]	1	1	1	0	0	3
Usman M. et al. [66]	1	1	0	0	1	3
Zhang G. et al. [68]	1	1	1	0	0	3
Deelstra S. et al. [43]	1	1	0	0	0	2
Hanssen GK. et al. [55]	1	1	0	0	0	2
Souza IS. et al. [64]	1	1	0	0	0	2
TOTAL	35	35	18	20	25	133

• The authors who have generated the most research are from Brazil, Germany, Spain, Canada and Sweden.

4. Related work

We have reviewed the works related to systematic studies of SLPE. We found a group of studies that made several efforts to systematically analyze various aspects of SPLE through systematic reviews of the literature (SLR) and systematic mapping studies (SMS). Table 18 shows a summary of these studies, which have been taken as a reference for our work, complementing them or responding to certain research gaps that resulted from the analysis of these studies.

These works mostly focus on specific domains or aspects of SPLE. Among these, Khurum and Gorschek [22] present a systematic review of domain analysis solutions published until 2007. This work aims to analyze the level of industrial application and the empirical validation of the proposed solutions in order to map the maturity in terms of industrial and utility application. The results of this research indicate that there are several solution for domain analysis in software product lines. However, the absence of qualitative and quantitative results of the empirical application or validation makes it difficult to evaluate the potential of the proposed solutions to be adopted by the industry. Likewise, We aim to identify the application and validation methods used in primary sources. Our approach addresses the empirical evaluations completed until November 2018 in all SPLE framework processes, including domain analysis.

Chen and Babar [23] reviewed and synthesized the status of the evaluation of variability management approaches published until 2007. Chen presented a detailed score card for the quality assessment of 18 empirical studies. Like our study, Chen evaluated the quality of variability management case studies based on the Runeson guidelines [29]. This study overlaps our work only on one study because only empirical evidence on variability management was analyzed, while our study covers all phases of SPLE.

El-Sharkawy et al. [24] aim to analyze metrics designed for variability models and code artifacts. In addition, they evaluated the metrics to provide a foundation for researchers to select suitable metrics. They identified the lack of contributions that analyze qualitative properties in the context of SPLE, emphasizing the need to improve the number of empirical validations. Our study also emphasizes the need to perform more empirical validations in SPLE that would allow researchers to base their contributions on works that have been validated in all SPLE topics.

Table 14

Number of papers per Journal or Conference.

Rank	No. Papers	JournalConference	EXP	CS	QUASI
1	15	International Systems & Software Product Line Conference	7	8	0
2	12	Journal of Systems and Software	2	9	1
3	12	Information and Software Technology	5	7	0
4	4	Empirical Software Engineering	3	1	0
5	3	Science of Computer Programming	1	2	0
6	3	International Symposium on Empirical Software Engineering and Measurement	2	1	0
7	2	Journal of Universal Computer Science	1	1	0
8	2	Software Quality Journal	1	1	0
9	2	Software - Practice and Experience	0	2	0
10	1	Journal of Biomedical Informatics	0	1	0
11	1	Requirements Engineering	0	1	0
12	1	User Modeling and User-Adapted Interaction	1	0	0
13	1	Expert Systems with Applications	1	0	0
14	1	Software and Systems Modeling	1	0	0
15	1	IET Software	1	0	0
16	1	International Conference on Software Engineering	0	1	0
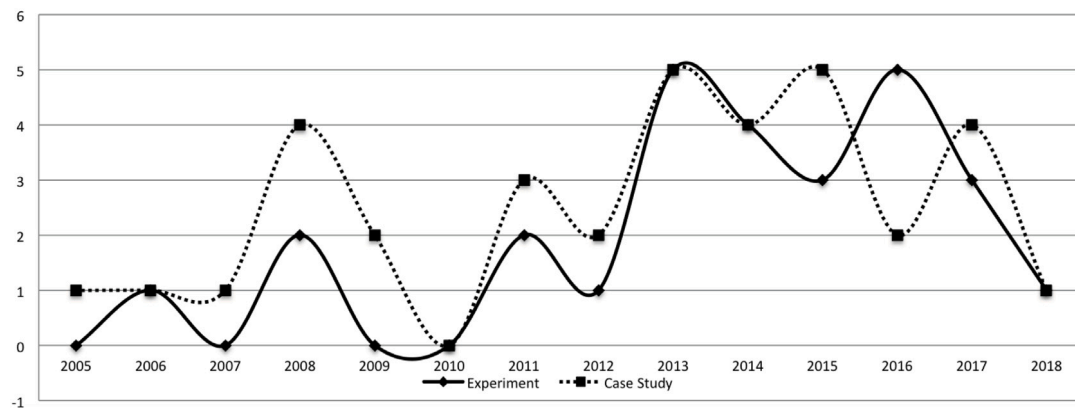
Fig. 9. Evolution of the studies according to the strategies.

Fig. 10. Classification of papers based on country.

Marques et al. [25] review the characteristics of the approaches that support the evolution of SPL and synthesize the evidence provided by the primary studies about the nature of their processes. They also analyze how the evolution of the SPLE is informed and validated. Among the reported findings, it is mentioned that the number of studies with some type of evaluation is growing in recent years. However,

Table 15
Empirical strategies classified by year.

Year	Case study	Experiment
2005	[42]	–
2006	[54]	[75]
2007	[36]	–
2008	[33] [55] [52] [65]	[70] [89]
2009	[43] [51]	–
2010	–	–
2011	[38] [61] [56]	[72] [81]
2012	[62][37]	[88]
2013	[49] [64] [50] [63] [59]	[34] [74] [35] [93] [78]
2014	[60] [46] [68] [40]	[84] [69] [90] [83]
2015	[47] [53] [39] [41] [45]	[77] [76] [80]
2016	[57] [48]	[71] [86] [87] [92] [91]
2017	[32] [44] [66] [67]	[85] [73] [79]
2018	[58]	[82]

they conclude that there is still a large number of articles without evaluation, since almost half of the studies of their SLR do not present any evaluation in their work. In our work, in addition to presenting how

SPLE approaches evolve, we also evaluate the quality of them based on guidelines from Runeson et al. [14] and Wohlin et al. [5].

Rabiser et al. [4] compare 142 papers but they goal to compare the topics addressed in industry and academy papers, ignoring other research questions that we study in this work. Furthermore, they limit the scope to the conference SPLC.

The survey from Berger et al. [30] focuses on the analysis of variability modeling in the industry with regard to the used strategy and the perceived values.

We also find an SMS that, similarly to the previous SLRs, is oriented to a specific area of the SPLE. This is the case of Laguna and Crespo [26], who study existing research on the reengineering of legacy systems in SPL and refactoring of SPL. One of the research questions sought to find out what types of validation studies are presented and up to what extension.

We also find two works that systematically address the revision of the literature in software product lines, through tertiary studies: the study by Marimuthu and Chandrasekaran [27] and the study by Raatikainen et al. [28].

The objective of the Marimuthu study is to carry out a systematic mapping of the existing systematic studies to inform the general vision

Table 16
Authors with more than two publications in empirical evidence according to the process.

#Papers	Author	PM	DRQ	DD	DE	DT	ARQ	AD	EE	AT	EXP	CE	QUASI
6	Eduardo Santana de Almeida	6	6	1	0	1	1	1	1	1	0	6	0
5	Silvio Romero de Lemos Meira	5	5	1	0	1	1	1	1	1	0	5	0
4	Paulo Anselmo da Mota Silveira Neto	4	4	1	0	1	1	1	1	1	0	4	0
4	Pádraig O'Learyd	4	4	1	0	1	1	1	1	1	0	4	0
3	Jan Bosch	3	3	2	0	1	2	2	1	1	0	3	0
3	Reinhartz-Berger, I.	0	3	2	0	0	0	0	0	0	3	0	0
3	Saake, G.	0	1	1	2	0	0	1	1	0	3	0	0
3	Sybren Deelstra	3	3	2	0	1	2	2	1	1	1	2	0
3	Sinnema Marco	3	3	2	0	1	2	2	1	1	1	2	0
2	Abrahão Silvia	1	1	1	0	0	0	1	0	0	1	1	0
2	Accioly, P.	0	1	0	0	1	0	0	0	1	2	0	0
2	Borba, P.	0	1	0	0	1	0	0	0	1	2	0	0
2	Bonifácio, R.	0	1	0	0	1	0	0	0	1	2	0	0
2	Capretz Luiz Fernando	1	0	1	0	0	0	1	0	0	1	1	0
2	Cetina, C.	0	0	1	0	1	0	1	0	0	1	1	0
2	Dachselt, R.	0	1	0	0	0	0	0	1	0	2	0	0
2	Díaz, J.	1	1	2	0	0	0	0	0	0	0	2	0
2	Faheem Ahmed	1	0	1	0	0	0	1	0	0	1	1	0
2	Feigenspan, J.	0	1	0	0	0	0	0	1	0	2	0	0
2	Figueiredo, E.	1	1	1	1	0	0	1	1	0	1	1	0
2	Figl, K.	0	2	1	0	0	0	0	0	0	2	0	0
2	Ivonei Freitas Da Silva	2	2	0	0	0	0	0	0	0	0	2	0
2	Garbajosa, J.	1	1	2	0	0	0	0	0	0	0	2	0
2	Gonzalez-Huerta Javier	1	1	1	0	0	0	1	0	0	1	1	0
2	Hanssen, G.K.	2	2	2	0	2	2	2	2	2	0	2	0
2	Insfran Emilio	1	1	1	0	0	0	1	0	0	1	1	0
2	Kästner, C.	0	1	0	0	0	0	0	1	0	2	0	0
2	Machado, I.	1	1	0	1	0	0	0	1	0	1	1	0
2	Martinez, J.	1	1	0	1	0	1	1	1	0	1	1	0
2	Øystein, H.	0	1	0	0	0	0	0	1	0	2	0	0
2	Papendieck, M.	0	1	0	0	0	0	0	1	0	2	0	0
2	Pérez Jennifer	1	1	2	0	0	0	0	0	0	0	2	0
2	Ribeiro, M.	0	1	0	1	0	1	1	0	0	2	0	0
2	Savolainen, J.	2	2	2	1	0	1	1	0	0	0	2	0
2	Schulze, M.	0	1	0	0	0	0	0	1	0	2	0	0
2	Ziadi, T.	1	1	0	1	0	1	1	1	0	1	1	0

Table 17

Authors with more than two publications in empirical evidences.

No. papers	Author	Country	Institutions
6	Eduardo Santana de Almeida	Brazil	Federal University of Bahia (UFBA)
5	Silvio Romero de Lemos Meira	Brazil	Federal University of Pernambuco (UFPE)
4	Paulo Anselmo da Mota Silveira Neto	Brazil	Reuse in Software Engineering (RiSE), Brazil
4	Pádraig O'Learyd	Ireland	Lero – The Irish Software Engineering Research Centre
3	Reinhartz-Berger Iris	Israel	University of Haifa
3	Jan Bosch	Sweden	Chalmers University of Technology
3	Saake Gunter	Germany	University of Magdeburg
3	Sybren Deelstra	The Netherlands	University of Groningen
3	Sinnema Marco	The Netherlands	University of Groningen
2	Kathrin Figl	Austria	Institute for Information Systems
2	Accioly Paola	Brazil	Federal University of Pernambuco
2	Borba Paulo	Brazil	Federal University of Pernambuco
2	Bonifácio Rodrigo	Brazil	University of Brasília
2	Figueiredo Eduardo	Brazil	Federal University of Minas Gerais
2	Ivonei Freitas da Silva	Brazil	Federal University of Pernambuco (UFPE)
2	Machado Ivan do Carmo	Brazil	Federal University of Bahia – UFBA
2	Márcio Ribeiro	Brazil	Federal University of Alagoas
2	Capretz Luiz Fernando	Canada	University of Western Ontario
2	Faheem Ahmed	Canada	University of Western Ontario
2	Savolainen Juha	Denmark	Danfoss Power Electronics A/S
2	Tewfik Ziadi	France	Sorbonne University UPMC
2	Dachselt Raimund	Germany	University of Magdeburg
2	Feigenspan Janet	Germany	University of Magdeburg
2	Kästner Christian	Germany	Philipps University Marburg
2	Papendieck Maria	Germany	University of Magdeburg
2	Schulze, M.	Germany	University of Magdeburg
2	Hanssen Geir K.	Norway	The Norwegian University of Science and Technology
2	Øystein Haugen	Norway	Østfold University College
2	Abrahão Silvia	Spain	Universitat Politècnica de València (UPV)
2	Cetina Carlos	Spain	Universidad San Jorge
2	Díaz Jessica	Spain	Technical University of Madrid-UPM
2	Garbajosa Juan	Spain	Technical University of Madrid-UPM
2	Gonzalez-Huerta Javier	Spain	Universitat Politècnica de València (UPV)
2	Insfran Emilio	Spain	Universitat Politècnica de València (UPV)
2	Martinez Jabier	Spain	University of Luxembourg
2	Pérez Jennifer	Spain	Technical University of Madrid-UPM

of the findings for researchers and professionals. 60 studies are listed in the area of software product lines. In the study of Marimuthu the first research question is finding out what kind of systematic study methods were considered. One of the conclusions of the study stated that many studies failed to evaluate the quality of the primary studies, and this was the main limitation of the systematic studies found. Our SLR focuses on finding out what empirical strategies have been used in SPLE and evaluating the quality of the empirical studies reported in

SPLE with human intervention. We infer that our study exceeds the limitation reported in the Marimuthu study because we address the evaluation of the primary studies.

Raatikainen et al. [28] study how 86 systematic reviews (SRs, i.e. SLR and SMS) analyze and use evidence-based results and identify how variability is modeled. One of the conclusions of the study was that the different types of SRs rarely take into account either the quality analyzed or the evidence found in the studies analyzed. This study also

Table 18

Comparison of related review papers.

Review work	Kind	Year	Title	Period	No. papers
Khurum et al. [22]	SLR (Systematic Literature Review)	2009	A systematic review of domain analysis solutions for product lines	–2007	89
Chen et al. [23]	SLR	2011	A systematic review of evaluation of variability management approaches in software product lines	1990–2007	97
El-sharkawy et al. [24]	SLR	2018	Metrics for analyzing variability and its implementation in software product lines : A systematic literature review	2007–2017	42
Marques et al. [25]	SLR	2018	Software product line evolution: A systematic literature review	1–2017	60
Laguna et al. [26]	SMS (Systematic Mapping Studie)	2013	A systematic mapping study on software product line evolution: From legacy system reengineering to product line refactoring	–2011	74
Marimuthu et al. [27]	SLR Terty	2017	Systematic Studies in Software Product Lines : A Tertiary Study	–2016	60
Raatikainen et al. [28]	SMS Terty	2018	Software product lines and variability modeling: A tertiary study	2007–2018	86
Rabiser et al. [4]	SMS	2018	A Study and Comparison of Industrial vs. Academic Software Product Line Research Published at SPLC	1997–2017	142
Our study	SLR	2019	Empirical Software product line engineering: A systematic literature review	2000–2018	62

ensured that many research methods are described briefly or are only mentioned by name. Even the studies analyzed are not always clear about their methods. They also mention that SRs are often quite specific for an emerging research area where solutions are proposed instead of research results with convincing empirical evidence that is worth analyzing. Therefore, several sub-themes may still lack an SR that is worth taking as a reference.

Our motivation to perform an SLR responds to the absence of a study that synthesizes and evaluates the quality of studies on empirical evidence in SPLE with human intervention. To our knowledge, there is no study with this topic, therefore we believe it is necessary to synthesize the studies with convincing empirical evidence of good quality so that researchers continue to make more contributions based on validated empirical evidence. We intend to reduce this gap through our SLR, addressing the analysis of the quality of the studies reported and showing the scope that empirical software engineering has had in the SPLE.

5. Conclusions

In this paper we presented a systematic review of the literature on empirical methods in SPLE with human intervention. We defined a protocol to carry out our research. We searched for documents from January 2000 to November 2018 and, after applying the inclusion and exclusion criteria, we selected 62 primary studies. We specified three research questions (Section 2.1.2). We defined a data extraction procedure to achieve the objectives of the SLR and answer the research questions.

We synthesized the current status of these primaries studies in order to (i) evaluate the quality determining the maturity of the method used in the primary studies; (ii) identify the scope of the empirical research in SPLE, i.e. which the processes of the SPLE framework attracted most attention of researchers; (iii) identify what kind of publications are proposed; (iv) when and where the articles were published; (v) who the authors and institutions are that currently have the know-how of empirical studies in SPLE.

We observed that the most used empirical strategy in SPLE topics is the case study with approximately 56%, while the rest of the primary studies used experimentation. Also, around 95% of the studies address aspects related to Domain Engineering, while there are other areas,

especially related to application engineering, that have received less attention. Similar results were obtained by Rabiser [4].

We also observed that there is a variety of topics studied. However, there is research interest in both academy and industry on issues related to variability management and modeling. A similar aspect is mentioned in the study presented by Rabiser [4].

The maturity of execution of the method used in the case studies was greater than the experiment and the information about evaluations or the examples presented in the documents are usually deficient. Comparable results are mentioned by Rabiser [4] "*The majority of academic research (61%) and also some industry research (11%) are not correctly validated (without evaluation or only artificial examples*". Nevertheless, as the Rabiser work also points out, the percentage of works with adequate empirical evaluation has been increasing over the years.

We consider that additional empirical studies should be performed with sufficient rigor to improve the body of evidence in SR within SPLE [31] because lack of justification in obtaining the sample of participants that will be used in the execution of the empirical strategy.

Due to the fact that there are few studies that address aspects related to application engineering, we consider that there is an opportunity to make more contributions in these processes. The cost of conducting studies based on experiments is not an easy task, as it must be planned, executed and analyzed properly, but nevertheless we believe it is necessary to perform more experiments in an industrial environment that allow validating the SPLE framework processes in practice. The objective of recommending the execution of more experiments in the industry is to validate and give more feedback to the SPLE framework adapted to the practice.

As results of this study, we remark that researchers in future contributions must carry out a more rigorous research design, with the aim of improving the quality of the method used. Researchers should be inspired by validated guidelines. We recommend Wholin [5] for experiments and Runeson [14] for case studies.

In addition, future research should pay special attention to the criteria that obtained the lowest score when evaluating the quality of the method used in the empirical strategy, to overcome the deficiencies presented in the studies evaluated. Regarding the research forum, researchers should consider sending their contributions (among others) to the Journal of System and Software, Information and Software Technology and Empirical Software Evidence since it has been verified

Table A.19
Search string and papers retrieved from each digital library.

Library	Search String	Result
ACM	+(''Software product line '' '' software product lines'' '' software product family'' '' software product families'')+(''empirical'' ''survey'' ''experiment'' ''case study'' ''experience report'')	659
IEEEXplore	((((''Software product line'' OR ''Software product lines'' OR ''Software product family'' OR ''Software product families'') AND (''Empirical'' OR ''Survey'' OR ''Experiment'' OR ''Case study'' OR ''Experience report'')))	283
Science Direct	for pub-date > 1999 and (''Software product line'' OR ''Software product lines'' OR ''Software product family'' OR ''Software product families'') AND (''Empirical'' OR ''Survey'' OR ''Experiment'' OR ''Case study'' OR ''Experience report'')[All Sources(Computer Science)]	613
Scopus	(TITLE-ABS-KEY (Software product line OR Software product lines OR Software product family OR Software product families) AND TITLE-ABS-KEY (Empirical OR Survey OR Experiment OR Case Study OR Experience report)) AND PUBYEAR > 1999 AND (LIMIT-TO (SUBAREA , ''COMP '')) AND (EXCLUDE (DOCTYPE , ''bk '') OR EXCLUDE (DOCTYPE , ''ch '') OR EXCLUDE (DOCTYPE , ''ip '') OR EXCLUDE (DOCTYPE , ''re '') OR EXCLUDE (DOCTYPE , ''ed '') OR EXCLUDE (DOCTYPE , ''er '')) AND (EXCLUDE (LANGUAGE , ''Portuguese '') OR EXCLUDE (LANGUAGE , ''French '') OR EXCLUDE (LANGUAGE , ''Japanese '') OR EXCLUDE (LANGUAGE , ''Chinese '') OR EXCLUDE (LANGUAGE , ''German '')) AND (EXCLUDE (DOCTYPE , ''cr''))	751

Table B.20
Data extraction properties.

N°	Property	Values	Description	R.Q.
1	Empirical strategy	- Experiment - Case Study - Quasy-Experiment - Survey	Empirical strategies given by [5]	RQ1.1.
2	Process Framework	- Product Management - Domain Requirement Engineering - Domain Design - Domain Realization - Domain Testing - Application Requirement Engineering - Application Design - Application Realization - Application Testing - Not specified.	Framework for Software Product Line Engineering of Pohl [2]	RQ1.2.
3	Research context	- Academic - Industrial	Type of context	RQ1.3.
4	Number of participants	- <= 20 participants - >20 or <= 100 participants - >= 100 participants - Not specified participants	Number of subjects	RQ1.3.
5	Sample Type	- Simple random sampling - Systematic sampling - Stratified random sampling - Random sampling by conglomerates - Convenience sampling - Quota sampling	Sample type	RQ1.3.
6	What it is evaluated	- Technique - Process - Tool	According to Basili [6]	RQ1.3.
7	Goal description	- Evaluates a proposal from the same paper - Evaluates an existing proposal - Is a replica of an experiment - Not specified	Goal description	RQ1.3.

Table B.21
Data extraction properties case studies.

N°	Property	Values	Description	R.Q.
1	Rationale for the Study	- Yes - No	The research clearly presents the reasons for conducting the study.	RQ2
2	Objective of the Study	- Yes - No	A statement is presented of what the researcher, and perhaps the industrial participants, hope to achieve as a result of conducting that study.	RQ2
3	Cases and Units of Analyses	- Yes - No	Is it mentioned in a general way or specifically what is being studied?	RQ2
4	Theoretical Framework	- Yes - No	Is it mentioned What is the theoretical frame of reference?	RQ2
5	Research Questions	- Yes - No	Is it mentioned What knowledge will be sought or expected to be discovered?	RQ2
6	Propositions and Hypotheses	- Propositions - Hypotheses - Both - Not Specified	Is it mentioned in the papers what particular (causal) relationships are to be investigated?	RQ2
7	Methods of Data Collection	- Yes - No	How will data be collected?	RQ2
8	Methods of Data Analysis	- Does not mention - Mentions but does not detail - Give few details - It gives more details - Details of the handicap	How will data be analyzed?	RQ2
9	Data Selection Strategy	- Yes - No	How will data be identified and selected? Who will be interviewed? What electronic data sources are available for use in the study?	RQ2
10	Quality Assurance, Validity, and Reliability	- Yes - No	How will the data collected be checked for quality? How will the analysis be checked for quality?	RQ2

that this type of contribution is within the scope of the topics of interest of this journal.

Materials

The results of searching in digital libraries and processing these searches were stored in the bibliography system Jabref. To access to these results, review https://doi.org/10.5281/zenodo.3860192 or https://zenodo.org/badge/latestdoi/189765534.

CRediT authorship contribution statement

Ana Eva Chacón-Luna: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - original draft, Writing - review & editing. **Antonio Manuel Gutiérrez:** Validation, Investigation, Formal analysis, Data curation, Writing - review & editing. **José A. Galindo:** Validation, Formal analysis, Data curation, Writing - review & editing, Visualization. **David Benavides:** Conceptualization, Data curation, Writing - review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Planning remarks

This section contains all the material and sources used in the data extraction.

- ACM: We use the advanced search in order to generate an initial set of papers. The results where stored in a bibtex file.
- IEEExplore: The search was restricted to the title and abstract of each publication, and the option "Full text and metadata" was enabled. The only format available for export from IEEExplore was .csv.
- Science Direct: First, we used the advanced search to generate an initial set of articles, and then we filtered the results obtained. We selected Computer Science as an area of knowledge. The results were stored in a bibtex file.
- Scopus: We restricted the search to titles of publications and abstracts, adding the words "TITLE-ABS" before the search terms. We also restricted the search to Computer Science by adding

Table B.22

Data extraction properties experiments.

N°	Property	Values	Description	R.Q.
1	Goal definition	- Yes - No	Object of study (what is studied?) and purpose (what is the intention?).	RQ2
2	Context selection	- Yes - No	The balance between making studies valid in a specific context or valid for the general domain of software engineering.	RQ2
3	Hypothesis formulation	- Yes - No	A hypothesis is stated formally.	RQ2
4	Variables selection	- Yes - No	Dependent and independent variables are chosen.	RQ2
5	Selection of subjects	- Yes - No	The selection of subjects or a sample of the population participating in the experiment is indicated.	RQ2
6	Choice of design type	- Yes - No	The design of an experiment describes how tests are organized and executed.	RQ2
7	Instrumentation	- Yes - No	The objects, guides or measuring instruments to be applied in the experiment are selected.	RQ2
8	Validity evaluation	- Yes - No	There is a section that demonstrates the validity of the experiment for example the conclusion, internal validity, construction or external validity.	RQ2
9	Preparation	- Yes - No	Before the experiment is executed, the participants are selected and informed, and the material is prepared as forms and / or tools.	RQ2
10	Execution	- Yes - No	It is described how the experiment is carried out: data collection (interviews, online forms?) or time period.	RQ2
11	Data validation	- Yes - No	The experimenter verifies that the data is reasonable and that it has been collected correctly.	RQ2
12	Descriptive statistics	- Yes - No	Statistics are presented on the numerical processing of the collected data set.	RQ2
13	Data set reduction	- Yes - No	Are statistical methods applied to detect outliers? Is the data set reduced if systematic errors are detected in the data?	RQ2
14	Hypothesis testing	- Yes - No	Some hypothesis test is performed to see if it is possible to reject the null hypothesis.	RQ2
15	Presentation & package	- Yes - No	There is a section devoted to the presentation of the results obtained from the execution of the experiment.	RQ2

"AND SUBAREA (" COMP ")". We exclude non-English papers by adding the words "EXCLUDE (LANGUAGE,"Portuguese", "French", "Japanese","Chinese","German")" and the types of documents that were not articles by adding "EXCLUDE (DOCTYPE, "bk", "ch", "ip", "re", "ed", "er"). The results were stored in a bibtex file.

Appendix B. Data extraction

Table B.20 shows the properties for the data extraction and, specifically, the third column lists the possible values that can be used to answer the RQ1 from the primary studies. Tables B.21 and B.22 show the forms used to extract the data related to the quality of the procedure for cases studies and experiment respectively. In both cases, the column Property mentions each element or verification criterion from which we seek to extract information from each primary study. The possible values for these elements are 1 if it meets the criterion or 0 if it does not meet it.

Table B.23

Characteristics that a case study report should have [15].

Characteristics	Characteristics for a case study
C1	To describe what the case study was about.
C2	To communicate a clear sense of the studied case.
C3	To provide a "history of the inquiry" so the reader can see what was done, by whom and how.
C4	To provide basic data in a focused form, so the reader can have confidence that the conclusions are reasonable.
C5	To articulate the researcher's conclusions and set them into a context they affect

References

[1] P. Clements, L. Northrop, Software Product Lines: Practices and Patterns,

Addison-Wesley Professional, 2001.

[2] K. Pohl, G. Böckle, F.J. van Der Linden, Software Product Line Engineering: Foundations, Principles and Techniques, Springer Science & Business Media, 2005.

[3] F.J. Van der Linden, K. Schmid, E. Rommes, Software Product Lines in Action: The Best Industrial Practice in Product Line Engineering, Springer Science & Business Media, 2007.

[4] R. Rabiser, K. Schmid, M. Becker, G. Botterweck, M. Galster, I. Groher, D. Weyns, A study and comparison of industrial vs. academic software product line research published at SPLC, in: Proceedings of the 22nd International Systems and Software Product Line Conference - Volume 1, SPLC 2018, Gothenburg, Sweden, September 10-14, 2018, 2018, pp. 14–24, http://dx.doi.org/10.1145/3233027.3233028.

[5] C. Wohlin, P. Runeson, M. Höst, M.C. Ohlsson, B. Regnell, A. Wesslén, Experimentation in Software Engineering, Springer Science & Business Media, 2012.

[6] V.R. Basili, R.W. Selby, D.H. Hutchens, Experimentation in software engineering, IEEE Trans. Softw. Eng. (7) (1986) 733–743.

[7] V.R. Basili, M.V. Zelkowitz, Empirical studies to build a science of computer science, Commun. ACM 50 (11) (2007) 33–37, http://dx.doi.org/10.1145/1297797.1297819.

[8] W.F. Tichy, Should computer scientists experiment more?, IEEE Comput. 31 (5) (1998) 32–40, http://dx.doi.org/10.1109/2.675631.

[9] C. Wohlin, M. Höst, K. Henningsson, Empirical research methods in software engineering, in: Empirical Methods and Studies in Software Engineering, Experiences from ESERNET, 2003, pp. 7–23, http://dx.doi.org/10.1007/978-3-540-45143-3_2.

[10] M.V. Zelkowitz, D.R. Wallace, Experimental models for validating technology, IEEE Comput. 31 (5) (1998) 23–31, http://dx.doi.org/10.1109/2.675630.

[11] B.A. Kitchenham, D. Budgen, P. Brereton, Evidence-Based Software Engineering and Systematic Reviews, vol. 4, CRC press, ISBN: 9781482228663, 2015, p. 426.

[12] B. Kitchenham, Procedures for performing systematic reviews, Keele, UK, Keele Univ. 33 (2004) (2004) 1–26.

[13] S. Keele, et al., Guidelines for Performing Systematic Literature Reviews in Software Engineering, Technical Report, Ver. 2.3 EBSE Technical Report. EBSE, 2007.

[14] P. Runeson, M. Höst, A. Rainer, B. Regnell, Case study research in software engineering, in: Guidelines and Examples, Wiley Online Library, 2012.

[15] C. Robson, Real World Research: A Resource for Social Scientists and Practitioner-Researchers, in: Regional Surveys of the World Series, Wiley, ISBN: 9780631213055, 2002, URL https://books.google.es/books?id=DkplMcAysFQC.

[16] E. Tacconelli, Systematic reviews: CRD's guidance for undertaking reviews in health care, Lancet Inf. Diseases 10 (4) (2010) 226.

[17] T. Dybå, B.A. Kitchenham, M. Jørgensen, Evidence-based software engineering for practitioners, IEEE Softw. 22 (1) (2005) 58–65, http://dx.doi.org/10.1109/MS.2005.6.

[18] J. Feigenspan, M. Schulze, M. Papendieck, C. Kästner, R. Dachselt, V. Köppen, M. Frisch, Using background colors to support program comprehension in software product lines, in: 15th Annual Conference on Evaluation Assessment in Software Engineering (EASE 2011), 2011, pp. 66–75.

[19] F. Shull, J. Singer, D.I. Sjøberg, Guide to Advanced Empirical Software Engineering, Springer, 2007.

[20] M. Cruz, B. Bernárdez, A. Durán, J.A. Galindo, A. Ruiz-Cortés, Replication of studies in empirical software engineering: A systematic mapping study, from 2013 to 2018, IEEE Access 8 (2020) 26773–26791.

[21] J. Chavarriaga, J.A. Hurtado, Second international workshop on experiences and empirical studies on software reuse (WEESR 2019), in: T. Berger, P. Collet, L. Duchien, T. Fogdal, P. Heymans, T. Kehrer, J. Martinez, R. Mazo, L. Montalvillo, C. Salinesi, X. Tërnava, T. Thüm, T. Ziadi (Eds.), Proceedings of the 23rd International Systems and Software Product Line Conference, SPLC 2019, Volume a, Paris, France, September 9-13, 2019, ACM, 2019, p. 44:1, http://dx.doi.org/10.1145/3336294.3342366.

[22] M. Khurum, T. Gorschek, A systematic review of domain analysis solutions for product lines, J. Syst. Softw. 82 (12) (2009) 1982–2003, http://dx.doi.org/10.1016/j.jss.2009.06.048.

[23] L. Chen, M.A. Babar, A systematic review of evaluation of variability management approaches in software product lines, Inf. Softw. Technol. 53 (4) (2011) 344–362, http://dx.doi.org/10.1016/j.infsof.2010.12.006.

[24] S. El-Sharkawy, N. Yamagishi-Eichler, K. Schmid, Metrics for analyzing variability and its implementation in software product lines: A systematic literature review, Inf. Softw. Technol. 106 (2019) 1–30, http://dx.doi.org/10.1016/j.infsof.2018.08.015.

[25] M. Marques, J. Simmonds, P.O. Rossel, M.C. Bastarrica, Software product line evolution: A systematic literature review, Inf. Softw. Technol. 105 (2019) 190–208, http://dx.doi.org/10.1016/j.infsof.2018.08.014.

[26] M.A. Laguna, Y. Crespo, A systematic mapping study on software product line evolution: From legacy system reengineering to product line refactoring, Sci. Comput. Program. 78 (8) (2013) 1010–1034, http://dx.doi.org/10.1016/j.scico.2012.05.003.

[27] C. Marimuthu, K. Chandrasekaran, Systematic studies in software product lines: A tertiary study, in: Proceedings of the 21st International Systems and Software Product Line Conference, SPLC 2017, Volume a, Sevilla, Spain, September 25-29, 2017, 2017, pp. 143–152, http://dx.doi.org/10.1145/3106195.3106212.

[28] M. Raatikainen, J. Tiihonen, T. Männistö, Software product lines and variability modeling: A tertiary study, J. Syst. Softw. 149 (2019) 485–510, http://dx.doi.org/10.1016/j.jss.2018.12.027.

[29] P. Runeson, M. Höst, Guidelines for conducting and reporting case study research in software engineering, Empir. Softw. Eng. 14 (2) (2009) 131–164, http://dx.doi.org/10.1007/s10664-008-9102-8.

[30] T. Berger, R. Rublack, D. Nair, J.M. Atlee, M. Becker, K. Czarnecki, A. Wasowski, A survey of variability modeling in industrial practice, in: The Seventh International Workshop on Variability Modelling of Software-Intensive Systems, VaMoS '13, Pisa , Italy, January 23 - 25, 2013, 2013, pp. 7:1–7:8, http://dx.doi.org/10.1145/2430502.2430513.

[31] V. Alves, N. Niu, C.F. Alves, G. Valença, Requirements engineering for software product lines: A systematic literature review, Inf. Softw. Technol. 52 (8) (2010) 806–820, http://dx.doi.org/10.1016/j.infsof.2010.03.014.

Primary sources

[32] J.F. Bastos, P.A. da Mota Silveira Neto, P. O'Leary, E.S. de Almeida, S.R. de Lemos Meira, Software product lines adoption in small organizations, J. Syst. Softw. 131 (2017) 112–128, http://dx.doi.org/10.1016/j.jss.2017.05.052.

[33] J. Kim, S. Park, V. Sugumaran, DRAMA: A framework for domain requirements analysis and modeling architectures in software product lines, J. Syst. Softw. 81 (1) (2008) 37–55, http://dx.doi.org/10.1016/j.jss.2007.04.011.

[34] J. Feigenspan, C. Kästner, S. Apel, J. Liebig, M. Schulze, R. Dachselt, M. Papendieck, T. Leich, G. Saake, Do background colors improve program comprehension in the #ifdef hell?, Empir. Softw. Eng. 18 (4) (2013) 699–745, http://dx.doi.org/10.1007/s10664-012-9208-x.

[35] A.K. Thurimella, B. Brügge, A mixed-method approach for the empirical evaluation of the issue-based variability modeling, J. Syst. Softw. 86 (7) (2013) 1831–1849, http://dx.doi.org/10.1016/j.jss.2013.01.038.

[36] F. Ahmed, L.F. Capretz, S.A. Sheikh, Institutionalization of software product line: An empirical investigation of key organizational factors, J. Syst. Softw. 80 (6) (2007) 836–849, http://dx.doi.org/10.1016/j.jss.2006.09.010.

[37] T. Berger, Variability modeling in the wild, in: International Software Product Line Conference, ISBN: 9781450310956, 2012, p. 287.

[38] A. Classen, Q. Boucher, P. Heymans, A text-based approach to feature modelling: Syntax and semantics of TVL, Sci. Comput. Program. 76 (12) (2011) 1130–1143, http://dx.doi.org/10.1016/j.scico.2010.10.005.

[39] G.C.B. Costa, R.M.M. Braga, J.M.N. David, F. Campos, A scientific software product line for the bioinformatics domain, J. Biomed. Inform. 56 (2015) 239–264, http://dx.doi.org/10.1016/j.jbi.2015.05.014.

[40] I.F. da Silva, P.A. da Mota Silveira Neto, P. O'Leary, E.S. de Almeida, S.R. de Lemos Meira, Software product line scoping and requirements engineering in a small and medium-sized enterprise: An industrial case study, J. Syst. Softw. 88 (2014) 189–206, http://dx.doi.org/10.1016/j.jss.2013.10.040.

[41] I.F. da Silva, P.A. da Mota Silveira Neto, P. O'Leary, E.S. de Almeida, S.R. de Lemos Meira, Using a multi-method approach to understand agile software product lines, Inf. Softw. Technol. 57 (2015) 527–542, http://dx.doi.org/10.1016/j.infsof.2014.06.004.

[42] S. Deelstra, M. Sinnema, J. Bosch, Product derivation in software product families: a case study, J. Syst. Softw. 74 (2) (2005) 173–194, http://dx.doi.org/10.1016/j.jss.2003.11.012.

[43] S. Deelstra, M. Sinnema, J. Bosch, Variability assessment in software product families, Inf. Softw. Technol. 51 (1) (2009) 195–218, http://dx.doi.org/10.1016/j.infsof.2008.04.002.

[44] S. Dey, S. Lee, REASSURE: Requirements elicitation for adaptive socio-technical systems using repertory grid, Inf. Softw. Technol. 87 (2017) 160–179, http://dx.doi.org/10.1016/j.infsof.2017.03.004.

[45] L.S. de Oliveira, P. O'Leary, E.S. de Almeida, S.R. de Lemos Meira, Product derivation in practice, Inf. Softw. Technol. 58 (2015) 319–337, http://dx.doi.org/10.1016/j.infsof.2014.07.004.

[46] J. Díaz, J. Pérez, J. Garbajosa, Agile product-line architecting in practice: A case study in smart grids, Inf. Softw. Technol. 56 (7) (2014) 727–748, http://dx.doi.org/10.1016/j.infsof.2014.01.014.

[47] J. Díaz, J. Pérez, J. Garbajosa, A model for tracing variability from features to product-line architectures: a case study in smart grids, Requir. Eng. 20 (3) (2015) 323–343, http://dx.doi.org/10.1007/s00766-014-0203-1.

[48] J. Echeverria, F. Perez, A. Abellanas, P. Ignacio, C. Cetina, O. Pastor, Evaluating bug-fixingin softwareproduct lines: an industrial case study, in: International Symposium on Empirical Software Engineering and Measurement, ISBN: 9781450344272, 2016.

[49] U. Eklund, H. Gustavsson, Architecting automotive product lines: Industrial practice, Sci. Comput. Program. 78 (12) (2013) 2347–2359, http://dx.doi.org/10.1016/j.scico.2012.06.008.

[50] E. Engström, P. Runeson, Test overlay in an emerging software product line - An industrial case study, Inf. Softw. Technol. 55 (3) (2013) 581–594, http://dx.doi.org/10.1016/j.infsof.2012.04.009.

[51] M. Eriksson, J. Börstler, K. Borg, Managing requirements specifications for product lines - An approach and industry case study, J. Syst. Softw. 82 (3) (2009) 435–447, http://dx.doi.org/10.1016/j.jss.2008.07.046.

[52] E. Figueiredo, N. Cacho, C. Sant'Anna, M. Monteiro, K. Uira, A. Garcia, S. Soares, F. Ferrari, S. Khan, F. Castor Filho, F. Dantas, Evolving software product lines with aspects: An empirical study on design stability, in: International Conference on Software Engineering, ACM, ISBN: 9781605580791, 2008, p. 882.

[53] M. Galster, P. Avgeriou, An industrial case study on variability handling in large enterprise software systems, Inf. Softw. Technol. 60 (2015) 16–31, http://dx.doi.org/10.1016/j.infsof.2014.12.003.

[54] D. Ganesan, D. Muthig, K. Yoshimura, Predicting return-on-investment for product line generations, in: International Software Product Line Conference, 2006.

[55] G.K. Hanssen, T.E. Fægri, Process fusion: An industrial case study on agile software product line engineering, J. Syst. Softw. 81 (6) (2008) 843–854, http://dx.doi.org/10.1016/j.jss.2007.10.025.

[56] G.K. Hanssen, Agile software product line engineering: enabling factors, Softw. Pract. Exper. 41 (8) (2011) 883–897, http://dx.doi.org/10.1002/spe.1064.

[57] H. Koziolek, T. Goldschmidt, T. de Gooijer, D. Domis, S. Sehestedt, T. Gamer, M. Aleksy, Assessing software product line potential: an exploratory industrial case study, Empir. Softw. Eng. 21 (2) (2016) 411–448, http://dx.doi.org/10.1007/s10664-014-9358-0.

[58] J. Martinez, X. Tërnava, T. Ziadi, Software product line extraction from variability-rich systems: The robocode case study, in: International Software Product Line Conference, vol. 1, Association for Computing Machinery, ISBN: 9781450363716, 2018, pp. 132–142, http://dx.doi.org/10.1145/3233027.3233038.

[59] V. Myllärniemi, J. Savolainen, T. Männistö, Performance variability in software product lines: A case study in the telecommunication domain, in: International Software Product Line Conference, ISBN: 9781450319683, 2013, pp. 32–41, http://dx.doi.org/10.1145/2491627.2491631.

[60] R.P. de Oliveira, D. Blanes, J. Gonzalez-Huerta, E. Insfrán, S. Abrahão, S. Cohen, E.S. de Almeida, Defining and validating a feature-driven requirements engineering approach, J. UCS 20 (5) (2014) 666–691, http://dx.doi.org/10.3217/jucs-020-05-0666.

[61] J. Bosch, P. Bosch-Sijtsema, Introducing agile customer-centered development in a legacy software product line, Softw. Pract. Exper. 41 (8) (2011) 871–882, http://dx.doi.org/10.1002/spe.1063.

[62] T. Patzke, M. Becker, M. Steffens, K. Sierszecki, Identifying improvement potential in evolving product line infrastructures: 3 case studies, in: International Software Product Line Conference, ACM, ISBN: 9781450310949, 2012.

[63] J. Rubin, K. Czarnecki, M. Chechik, Managing cloned variants: A Framework and experience, in: International Software Product Line Conference, 2013, p. 286.

[64] I.S. Souza, G.S. da Silva Gomes, P.A. da Mota Silveira Neto, I. do Carmo Machado, E.S. de Almeida, S.R. de Lemos Meira, Evidence of software inspection on feature specification for software product lines, J. Syst. Softw. 86 (5) (2013) 1172–1190, http://dx.doi.org/10.1016/j.jss.2012.11.044.

[65] A.K. Thurimella, B. Bruegge, O. Creighton, Identifying and exploiting the similarities between rationale management and variability management, in: International Software Product Line Conference, IEEE Computer Society, ISBN: 9780769533032, 2008, pp. 99–108, http://dx.doi.org/10.1109/SPLC.2008.14.

[66] M. Usman, M.Z. Iqbal, M.U. Khan, A product-line model-driven engineering approach for generating feature-based mobile applications, J. Syst. Softw. 123 (2017) 1–32, http://dx.doi.org/10.1016/j.jss.2016.09.049.

[67] D. Wille, K. Wehling, C. Seidl, M. Pluchator, I. Schaefer, Variability mining of technical architectures, in: International Software Product Line Conference, vol. 1, Association for Computing Machinery, ISBN: 9781450352215, 2017, pp. 39–48, http://dx.doi.org/10.1145/3106195.3106202.

[68] G. Zhang, H. Ye, Y. Lin, Quality attribute modeling and quality aware product configuration in software product lines, Softw. Qual. J. 22 (3) (2014) 365–401, http://dx.doi.org/10.1007/s11219-013-9197-z.

[69] P.R.G. Accioly, P. Borba, R. Bonifácio, Controlled experiments comparing black-box testing strategies for software product lines, J. UCS 20 (5) (2014) 615–639, http://dx.doi.org/10.3217/jucs-020-05-0615.

[70] F. Ahmed, L.F. Capretz, The software product line architecture: An empirical investigation of key process activities, Inf. Softw. Technol. 50 (11) (2008) 1098–1113, http://dx.doi.org/10.1016/j.infsof.2007.10.013.

[71] M. Asadi, S. Soltani, D. Gasevic, M. Hatala, The effects of visualization and interaction techniques on feature model configuration, Empir. Softw. Eng. 21 (4) (2016) 1706–1743, http://dx.doi.org/10.1007/s10664-014-9353-5.

[72] E. Bagheri, D. Gasevic, Assessing the maintainability of software product line feature models using structural metrics, Softw. Qual. J. 19 (3) (2011) 579–612, http://dx.doi.org/10.1007/s11219-010-9127-2.

[73] R. Bonifácio, P. Borba, C. Ferraz, P.R.G. Accioly, Empirical assessment of two approaches for specifying software product line use case scenarios, Softw. Syst. Model. 16 (1) (2017) 97–123, http://dx.doi.org/10.1007/s10270-015-0471-3.

[74] C. Cetina, P. Giner, J. Fons, V. Pelechano, Prototyping dynamic software product lines to evaluate run-time reconfigurations, Sci. Comput. Program. 78 (12) (2013) 2399–2413, http://dx.doi.org/10.1016/j.scico.2012.06.007.

[75] C. Denger, R. Kolb, Testing and inspecting reusable product line components: First empirical results, in: International Symposium on Empirical Software Engineering and Measurement, 2006.

[76] D. Dermeval, T. Tenório, I. Ibert, A. Silva, S. Isotani, M. Ribeiro, Expert systems with applications ontology-based feature modeling : An empirical study in changing scenarios, Expert Syst. Appl. (ISSN: 0957-4174) 42 (11) (2015) 4950–4964, http://dx.doi.org/10.1016/j.eswa.2015.02.020.

[77] J. Gonzalez-Huerta, E. Insfrán, S.M. Abrahão, G. Scanniello, Validating a model-driven software architecture evaluation and improvement method: A family of experiments, Inf. Softw. Technol. 57 (2015) 405–429, http://dx.doi.org/10.1016/j.infsof.2014.05.018.

[78] V. Guana, D. Correal, Improving software product line configuration: A quality attribute-driven approach, Inf. Softw. Technol. 55 (3) (2013) 541–562, http://dx.doi.org/10.1016/j.infsof.2012.09.007.

[79] Y. Liu, L. Liu, H. Liu, X. Wang, H. Yang, Mining domain knowledge from app descriptions, J. Syst. Softw. 133 (2017) 126–144, http://dx.doi.org/10.1016/j.jss.2017.08.024.

[80] J. Martinez, T. Ziadi, T.F. Bissyandé, J. Klein, Y. Le Traon, Bottom-up adoption of software product lines - A generic and extensible approach, in: International Software Product Line Conference, 20-24-July-2015, Association for Computing Machinery, ISBN: 9781450336130, 2015, pp. 101–110, http://dx.doi.org/10.1145/2791060.2791086.

[81] B. Michalik, D. Weyns, N. Boucké, A. Helleboogh, Supporting online updates of software product lines: A controlled experiment, in: International Symposium on Empirical Software Engineering and Measurement, 2011, pp. 187–196.

[82] J.A. Pereira, S. Schulze, E. Figueiredo, G. Saake, N-dimensional tensor factorization for self-configuration of software product lines at runtime, in: International Software Product Line Conference, vol. 1, Association for Computing Machinery, ISBN: 9781450363716, 2018, pp. 87–97, http://dx.doi.org/10.1145/3233027.3233039.

[83] I. Reinhartz-Berger, K. Figl, Comprehensibility of orthogonal variability modeling languages: The cases of CVL and OVM, in: International Software Product Line Conference, vol. 1, Association for Computing Machinery, ISBN: 9781450327404, 2014, pp. 42–51, http://dx.doi.org/10.1145/2648511.2648516.

[84] I. Reinhartz-Berger, A. Sturm, Comprehensibility of UML-based software product line specifications - a controlled experiment, Empir. Softw. Eng. 19 (3) (2014) 678–713, http://dx.doi.org/10.1007/s10664-012-9234-8.

[85] I. Reinhartz-Berger, K. Figl, Ø. Haugen, Investigating styles in variability modeling: Hierarchical vs. constrained styles, Inf. Softw. Technol. 87 (2017) 81–102, http://dx.doi.org/10.1016/j.infsof.2017.01.012.

[86] M. Saeed, F. Saleh, S. Al-Insaif, M. El-Attar, Empirical validating the cognitive effectiveness of a new feature diagrams visual syntax, Inf. Softw. Technol. 71 (2016) 1–26, http://dx.doi.org/10.1016/j.infsof.2015.10.012.

[87] A.R. Santos, I. Do Carmo Machado, E.S. De Almeida, RiPLE-HC: JavaScript systems meets SPL composition, in: International Software Product Line Conference, 16-23-September-2016, Association for Computing Machinery, ISBN: 9781450340502, 2016, pp. 154–162, http://dx.doi.org/10.1145/2934466.2934486.

[88] J. Feigenspan, M. Schulze, M. Papendieck, C. Kästner, R. Dachselt, V. Köppen, M. Frisch, G. Saake, Supporting program comprehension in large preprocessor-based software product lines, IET Softw. 6 (6) (2012) 488–501, http://dx.doi.org/10.1049/iet-sen.2011.0172.

[89] M. Sinnema, S. Deelstra, Industrial validation of COVAMOF, J. Syst. Softw. 81 (4) (2008) 584–600, http://dx.doi.org/10.1016/j.jss.2007.06.002.

[90] J. Stein, I. Nunes, E. Cirilo, Preference-based feature model configuration with multiple stakeholders, in: International Software Product Line Conference, vol. 1, Association for Computing Machinery, ISBN: 9781450327404, 2014, pp. 132–141, http://dx.doi.org/10.1145/2648511.2648525.

[91] T. Thüm, M. Ribeiro, R. Schröter, J. Siegmund, F. Dalton, Product-line maintenance with emergent contract interfaces, in: International Software Product Line Conference, 16-23-September-2016, Association for Computing Machinery, ISBN: 9781450340502, 2016, pp. 134–143, http://dx.doi.org/10.1145/2934466.2934471.

[92] A. Vasilevskiy, F. Chauvel, Ø. Haugen, Toward robust product realisation in software product lines, in: International Software Product Line Conference, 16-23-September-2016, Association for Computing Machinery, ISBN: 9781450340502, 2016, pp. 184–193, http://dx.doi.org/10.1145/2934466.2934484.

[93] Y. Wang, A. Kobsa, A PLA-based privacy-enhancing user modeling framework and its evaluation, User Model. User-Adapt. Interact. 23 (1) (2013) 41–82, http://dx.doi.org/10.1007/s11257-011-9114-8.