Multi Software Product Lines: A Systematic Mapping Study

Pasquale Ardimento[a], Nicola Boffoli[b] and Giuseppe Superbo
Computer Science Department, University of Bari Aldo Moro, Via Orabona, 4, 70125, Bari, Italy

Keywords: Multi Software Product Line, Software Product Line, Software Engineering, Mapping Study, Variability Management.

Abstract: Even if Software Product Line (SPL) is an established technique in software engineering, there are several limitations of its use. This is mainly caused by the exponential increase of software systems complexity and by the high pace of software and hardware evolution. Many of these limitations have been studied and faced by applying the Multi Software Product Line (MSPL), an extension of the SPL. MSPL is an emerging and novel technique based on using more than one SPL to derive a functional product system. This paper aims to characterize the state of the art of MSPL, the main goal is to highlight the reached achievement in this field and then discuss about the open issues or the not covered aspects of this approach. In order to make an overall analysis inside the research community, a well-defined method of systematic mapping is applied in order to classify, in a proper scheme, every paper strictly related to this topic. These classified results could give a valid hint about lacks which should be investigated further.

1 INTRODUCTION

Nowadays Software Product Line (SPL) is a common development approach which is used in many context of Software Engineering. As the term of concept exploits, this technique consists in applying all the traditional product line methods in the software development sector. So as there are several selectable components to build a car, also during the development of a software system it is possible to choose existent features/modules from a less specific platform (Pohl et al., 2005). In this way it is possible to build several customized systems for a very specific context without any additional effort in the development stage. Despite all the advantages in costs and time, there are some critical limitations which are caused by the complexity of software market environment. The high pace of software and hardware evolution, the frequent change of requirements and the large size of systems make configuration and maintenance of a production platform really complex. For this reason after many year of research, an extension of this concept was identified as potential solution of these limitations: Multiple Software Product Line (MSPL) (Krueger, 2006). This derived development strategy is based in building new systems by selecting features/assets from different software product lines without altering any functionality of each SPL platform, which are still independent between each others. This can be applied in many contexts as often totally disconnected fields can share several features/assets or they can share some technology features in their core functionalities (Marinho et al., 2013). However as many limitations of SPL are solved with the introduction of this new approach, many other issues are introduced with the use of MSPL. These problems are mainly caused by the lack of real application as a lot of open issues still make the management of this technique really hard. For example working with software product lines made by different organisations or companies it is really complex as they can rely on different architecture or they can be designed in a different way. For this reason the degree of reuse depends on how many SPLs are used to derive the final product. Additionally there are no tools available for the management of a MSPL as most of SPLs tool are extended with MSPL functionalities which does not guarantee a complete and satisfying management of the MSPL. Developers are so forced to adapt and to interface SPLs tools to achieve a satisfying level of usability and maintenance with a MSPL. Finally it is possible to notice that automating this entire process is still an open issue as it is really complex to make decision upon the selection order of

[a] https://orcid.org/0000-0001-6134-2993
[b] https://orcid.org/0000-0001-9899-6747

Ardimento, P., Boffoli, N. and Superbo, G.
Multi Software Product Lines: A Systematic Mapping Study.
DOI: 10.5220/0009469804700476
In *Proceedings of the 15th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE 2020)*, pages 470-476
ISBN: 978-989-758-421-3

the components available from different SPLs.

The goal of this paper is to achieve an overview of the field of MSPL and the authors, at this aim, choose to accomplish a systematic mapping of the existent literature (Kitchenham et al., 2010). In general, a systematic mapping study structures the type of research, reports and results that have been published by categorizing them and yields a visual summary, systematic map, of its results. In this work, the map built and the related discussion of the findings aim to support understanding what has been addressed by the MSPL community for the MSPL domain and to suggest future research guidelines in this field.

The reminder of this paper is structured as follows: section 2 describes the research methodology used for the mapping study carried out in this work, from establishing the research questions to the resulting systematic map. After this, in Section 3 there is the analysis and discussion of the results extracted during the mapping study. Section 4 discusses the threats to validity of our study. Finally, Section 5 summarizes the main contribution of this paper and discusses some directions that could be investigated by the research community.

2 RESEARCH METHODOLOGY

To accomplish all the prefixed goals, the authors have adopted Kitchenham's approach for performing a systematic mapping study (Kitchenham et al., 2008). This method is helpful to reach the goals of the paper if there is enough available content for a specific topic: all the final results exploit which areas need to be still explored and studied. Figure 1 shows all the steps of the systematic mapping process.

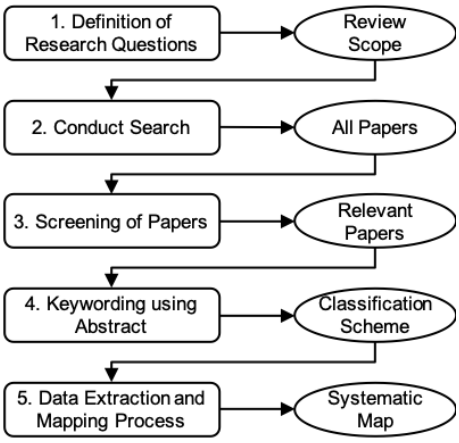

Figure 1: Mapping Process.

In the follows of the section, steps from 1 to 5 are described in details and applied to the mapping of the MSPL literature.

2.1 Definition of Research Questions

In the first step, the scope of this study is characterized. According to these initial suppositions it is possible to develop a schematic guideline of which questions are going to characterize the research:

- *RQ1*: Which MSPL life-cycle step affecting variability is covered by the research community?
- *RQ2*: Which of the found topics are less covered and could be considered still "open issues"?

From the results extracted through the systematic mapping, we can infer which problems are not tackled inside the research works and require more solution proposals. Moreover, after having populated the classification scheme, different facets can also be combined to answer more specific research questions:

- *RQ3*: Concerning each topic, which level of empirical evidence is reached by researchers's solutions for each specific problem?
- *RQ4*: Concerning each topic, which type of innovation (metric, tool, model, method, process) is needed to recognize an issue as "closed"?

2.2 Conduct Search for Primary Studies

After introducing the systematic mapping method in general, here it will be applied to the paper context in order to extract useful information from the research community. First of all it is necessary to choose a subset of search string according to two of Kitchenham's criteria (Kitchenham et al., 2008):

1. Population: Multiple software product line, software product line, hierarchical software product line, nested software product line;
2. Intervention: variability management.

Following this criteria, it is possible to obtain a very specific search string. However, since a search bar of a Web browser has a limit of 150 characters, the search string was split in 3 sub-strings:

1. "Multi product line" OR "Multi Product lines" OR "Multi product family" OR "Multi product families" OR "Multi software product line" OR "Multi software product line" OR "Multi software product lines" OR "Multi software product family" returned 595 results from 2010.
2. "Multi software product families" OR "Multiple product line" OR "Multiple product lines" OR

"Multiple product family" OR "Multiple product families" OR "Multiple software product line" OR "Multiple software product lines" returned 2170 results from 2010.

3. "Multiple software product family" OR "Multiple software product families" OR "Product line of product lines" OR "Nested product lines" OR "Hierarchical product line" OR "Hierarchical product lines" returned 93 results from 2010.

Basically the search string should select any paper regarding the variability management aspect of MSPL. All the initial papers were selected by searching any content in scientific virtual libraries. The previous research tokens were applied on one of the richest Web search engine for scientific papers, technical reports, theses and books: Google Scholar. This free access powerful tool let anyone to customize the research in terms of time and number citations. In this way it is possible to select only the relevant content in a specific interval of time. In the specific case of this paper, time interval for the research is set to 2010 to current days as the MSLP concept was defined only in these last years of research. Once the search engine returns all the results, it is possible to select from them only the ones coming from sources like ACM, IEEE, Elsevier, Springer. Additionally it is also possible to select content from official conferences on that specific topic. In this case two of the most related conference on the paper topic are SPLC (Software Product Line Conferences), which started in the early 2000s, and VaMoS (Variability Modelling of Software-Intensive Systems). During these conferences several workshops are held. In SPLC 2013 the organization held a workshop upon MSLP called MultiPLE2013 from which it is possible to select valid content. Unfortunately even if during SPLC 2014, MultiPLE2014 was again scheduled, it was cancelled due several reasons. During the analysis all the content coming from this workshop will be considered part of SPLC as this workshop was held during this conference.

2.3 Screening of Papers

To reduce the corpus and enable this study's reproducing, the following inclusion criteria and exclusion criteria were established.

Inclusion criteria:

- Peer-reviewed studies published in journals, conferences, and workshops.
- Studies are accessible electronically.
- From title, abstract, and keywords we can deduce if the paper focuses on contributing MSPL.

Exclusion criteria:

- Studies not available in English.
- Studies to be excluded from a systematic peer-review, such as websites, books, slides, teasers, curricula, short papers and editorials of less than two pages.
- Studies where MSPL is mentioned as future application, related work, or broad context only.

We applied both the criteria to the following metadata: abstract, title and keywords. This step aims at eliminating only the publications clearly not within our study's scope and publications failing on formal requirements (such as being available in English). It was necessary to discard most of the documents from the results of the queries as most of them are not related with the paper main topic. Anyway, all studies with any doubt about their relevance to the study topic were included to be evaluated in the next step on the base of their full text. At the end of this step we obtained the results shown in table 1 where the main sources that contribute to the paper goal are listed.

Table 1: Resulting Papers.

	PAPERS (2010-2019)
ACM	6
Elsevier	5
IEEE	9
SPLC	11
Springer	6
VaMoS	5
Others	7
TOTAL	49

2.4 Keywording using Abstract

After filtering the relevant content from the entire research community, the authors read all the collected paper to make a double check if they are really relevant to the paper goal and to classify the relevant ones. The classification scheme was made by a set of categories representing the underlying population. To exploit all the covered aspects of MSPL by the research community, the classification scheme was composed of three aspects which are MSPL Topics, Evidence and Innovation.

Facet 1: Topics of MSPL. Nevertheless as MSPL differs from SPL, they share the same core idea: variability management. This crucial aspect means, as the term suggests, managing variability in software final artifacts. This occurs obviously during the entire lifecycle of the product line as it is really important to track every change in variability points in or-

der to avoid obsolescence inside the SPL (more details in (Clements and Northrop, 2002)). This task is even harder in MSPL as more SPLs are involved in the variability management. These crucial life-cycle steps of SPL are a good guideline to analyze how MSPL can be organized in the same way, therefore such steps have been chosen as MSPL variability topics because they could be considered a strong landmark for the classification of all existent papers related to MSPL. This facet aims to address the discussion of RQ1, RQ2 and it is structured as follows:

- *Domain Engineering*: in the early stages of the development of a PL, as in every software life-cycle, developers collect information from stakeholders regarding the domain. In this way it is possible to target which are the common and variant assets, along with their feature modeling and evaluation of the product line scoping (a.k.a. domain analysis);
- *Product Derivation*: once an abstract view of the domain is established thanks to the previous step, it is time to focus on variability in software architecture modeling and construction. This means that in order to establish a working PL, it is mandatory to design different compositions of common and variant assets according to the requirements collected previously (a.k.a. architecture design);
- *Product Configuration*: as it occurs in every development process, also in the PL life-cycle there is a step where each final product is configured by selecting certain parameters and variables according to each operative context (a.k.a. implementation);
- *Product Line V&V*: in order to guarantee a good level of quality of the final artifact, developers follow several rules and constraints to design optimal test cases which are always based upon the operative context of the final product (a.k.a. verification and validation);
- *Product Line Maintenance*: once the product is released and finally executed, as it happens for every software product, there is the most expensive phase of a product life-cycle: maintenance. In this context maintainers focus on keeping track of versions, dealing with changes in requirements at all abstraction levels and correct any issue related with the use of any final artifact (a.k.a. evolution and change).

Facet 2: Evidence. Evidence facet aims to support the discussion about the RQ3. We adopted the existing classification defined by (Wieringa et al., 2006), in this way it is possible to categorize all the papers as follows (each paper was classified to belong to exactly one research type):

- *Early Research Advanced (ERA) Paper*: a paper with a new way of looking at existing issues, including surveys, mapping studies and systematic reviews.
- *Solution Proposal Paper*: a novel solution for a problem is proposed or it can be a significant extension of an existing technique. The potential benefits and the applicability of the solution can be explained by means of an illustrative example.
- *Validation Paper*: novel solutions are investigated (for example, by means of a controlled experiment in a laboratory), hypotheses are stated and statistics are shown but they are not evaluated yet in a real world project.
- *Evaluation Paper*: solutions are applied in a real world project and an empirical evaluation is conducted to prove their efficacy in a rigorous way, based on measures collected in a real context.
- *Experience Report*: the paper simply reports the practical experience whether some technique or methodology is good or bad inside a real context (industrial or laboratory).

Facet 3: Innovation. To address RQ4, we classified the publications according to the innovative solution proposed to solve a specific open issue, such classification is inspired by (Petersen et al., 2008):

- *Overview*: which includes surveys, mapping studies and systematic reviews;
- *Process/method*: which includes a workflow of activities, rules or procedures about how things should be done (i.e. how to recognize common and variant assets);
- *Model*: defining a formal abstraction of variability details and related semantics and notations;
- *Metric*: describing a metric plan to measure variability at different levels;
- *Tool*: developing a software tool that support different aspects of MSPL variability.

2.5 Data Extraction and Mapping Process

As soon as all the related papers were classified properly, a systematic map is built (figure 2). Thanks to this plot, it is now possible to have statistics about each category content and sketch an overview of how the research community covered MSLP topic. By

looking at the schematic map it is possible to see which are the most covered subtopics and the ones that lack of content.

More precisely all the papers characteristics are categorized in a bubble plot. This plot is made up of two two-dimensional (x-y) scatter-plots with frequency bubbles for each intersection. The first one connects all the innovation facet with all the steps of MSPL process. Meanwhile in the second plot there is the conjunction between evidence facet and steps of MSPL process. The bubble size is directly proportional to the number of papers that are characterized by the intersection of any two categories. Finally it is important to point out that the total number of the frequencies is not the same as the total number of the papers as they may have different facet.

3 MAIN FINDINGS

Supported from the systematic map (figure 2), the final results of the systematic mapping process can be analyzed and discussed in this section in order to address the research questions reported in 2.2.

RQ1: "Most Covered MSPL Steps". By looking at "MSPL topic" of the systematic map it is possible to notice that: 27 papers (55.10%) focused on *domain engineering*; 30 papers (61.22%) focused on *product derivation*; 23 papers (46.93%) focused on *product configuration*; 9 papers (18.36%) focused on *product line V&V*; 6 papers (12.24%) focused on *product line maintenance*. As it is possible to see from the statistics *domain engineering* and *product derivation* are the most covered topic in this paper research. This means that researchers started of course to mine the topic by starting from the first crucial steps of MSPL development. According to the map in figure 2 many researchers provided their own innovative contribution in these two topic mainly in terms of *process/method* or *model*. Additionally, it is possible to observe from the corpus references (see **??**) that these two topics are often covered at the same time, also *product configuration* is covered well enough and with this topic it is possible to find a strong relation to *product derivation* as they are strictly linked and studied at the same time. So, we can realize that in general, steps that in the lifecycle are strongly connected to each other are often dealt simultaneously in the same paper.

RQ2: "Less Covered MSPL Topic". What is not enough covered by activity research, is for sure the last two phases of the development cycle: *product line V&V* (9 papers, 18.36%) and *product line maintenance* (6 papers, 12.24%). Regardless of the innovation and the evidence facets, *V&V* and *maintenance* remain the current open issues in the research field, this might suggest that testing, fixing and evolving a MSPL could be valid topics that can be investigated more deeply. This lack of content is probably related to the challenges of *domain engineering* and *product derivation* that are still being discussed. For this reason it is possible to recognize a priority pattern among the MSPL topics: once *domain engineering* and *product derivation* are going to have standardized approaches, research community will probably focus their attention on the maintenance and testing side (in fact, such papers have been published only in the last five years).

RQ3: "Level of Empirical Evidence". Here there is an average distribution between three of five facets as *solution proposal*, *validation papers* and *evaluation papers* are nearly equally distributed: 14 papers (28.57%) were *solution proposal*; 16 papers (32.65%) were *validation papers*; 13 papers (26.53%) were *evaluation papers*. This means that researchers were focused on testing approaches, trying solving problems and even conducting empirical validation as 13 papers were *evaluation papers*. On the other hand it was possible to retrieve only 2 *experience reports* (4.08%) and 4 *ERA papers* (8.16%): the cause of the lack of the first kind of papers is related maybe to the lack of real application of MSPL. Meanwhile for the *ERA papers* the lack of content is related to the lack of necessity in writing this typology of paper about MSPL field (number of papers to investigate in this field was low).

RQ4: "Innovation Needed". Speaking about innovation facet 35 papers (71.42%) provide a new *process/method*; 16 papers (32.65%) provide a new *model*; 12 papers (24.48%) provide a new *tool*; 9 papers (18.36%) provide a new *metric*; 3 papers (6.12%) provide an *overview on the topic*. Many researchers have been focused on proposing new *processes/methods* to solve all the open issues regarding MSPLs. Despite of this, a third of the papers contributed also with a new *model* that helped this young development approach to be recognized and used in the software industry. Additionally many *tools* and extensions of theme were proposed to make the use of this approach less complex. Nonetheless there is a low quantity of *metrics* paper as it is still complex to measure precisely the variability at different levels. Finally *overview* paper are only two for the same reason of the low quantity of ERA papers.

In conclusion, in the last decade many new approaches were designed and even used in real use cases. By looking at these new approaches, it is very easy to observe that most of them are strictly related

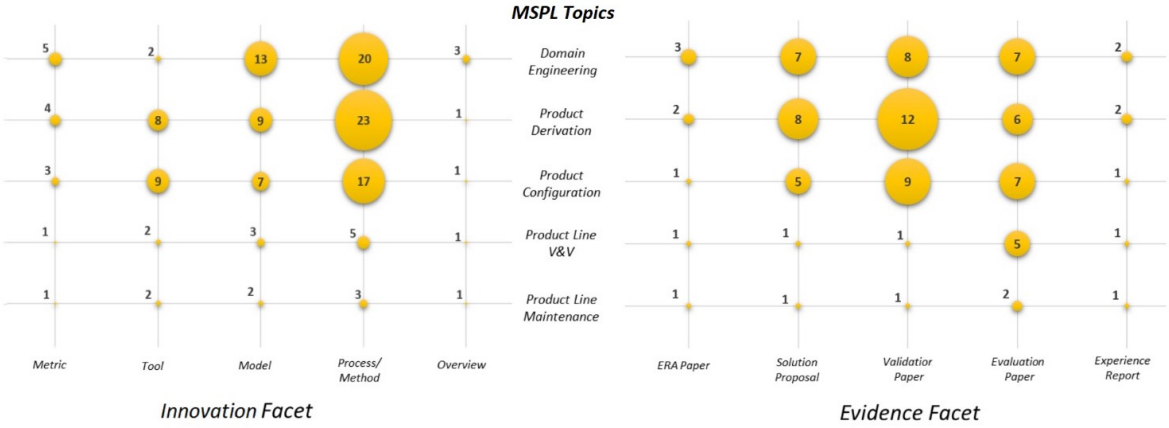

Figure 2: Topic coverage according to the Innovation and Evidence facet.

to the most common topics among the available papers on MSPLs: "Processes/Methods" regarding the innovation facet and "Validation" regarding validation paper. Hence it is obvious that there is a strong trend emerging from the research community: finding a practical and pragmatical use of the MSPL. In addition to this, it is also clear that Delta-oriented Multi Software Product Lines (Damiani et al., 2014), INVAR approach "plug-and-play" for variability models (Dhungana et al., 2011) and AGILE product line engineering applied to MSPLs (Hayashi and Aoyama, 2018) are example of practices/processes that have been tested on real use cases simulations. Most of the new approaches focuses on the crucial steps of a solid MSPL: Core Assets Development and Product development.This underlines that a successful MSPL can be achieved only if a good basis is established.

4 THREATS TO VALIDITY

To judge the quality of our work it is necessary to discuss the following threats to the validity of the study.

Construct Validity. The findings presented in section 3 are valid only for our sample of papers. Hence, we ensured the inclusion of as many relevant papers as possible. To achieve this, the research tokens were applied to the Google Scholar digital library in order to capture venues not published at the other libraries. Then, we applied the research tokens also to the main repositories of resources dealing with the selected topic. Furthermore, we did not restrict our search to papers mentioning "Multi Software Product Lines" explicitly but also included terms closely related to MSPL, without narrowing it to the exact terms. This enabled us to capture related publications without focusing on the very specific, partly ambiguous, modeling terminology. Another threat avoided regards the definition of the inclusion/exclusion criteria in the screening activity. We decided to consider the title, abstract, and keywords metadata. We also decided to include all papers we were uncertain in order to avoid the exclusion of relevant publications due to the lack of investigation. In the next step, first, all the papers were read and, then, inclusion/exclusion was decided. Finally, to improve construct validity in the future we should consider also the papers that do not use the term "multiple software product line" even though they are clearly discussing MSPL domain modeling techniques, for example among the main works of this type there are (Bagheri et al., 2011) (Hubaux et al., 2013) (Urli et al., 2014).

Internal Validity. To mitigate the threat related to the data extraction performed for screening (inclusion and exclusion criteria) and classification, the authors decided to discuss the problematic papers among them. The classification was performed in blocks of at most 45 minutes broken up by at least 15-minute breaks in order to avoid the threat arising from fatigue. The adopted classification scheme comprises three facets that correspond to the four defined research questions.

Conclusion Validity. Regarding the conclusions, we have discussed various issues that could lead to wrong conclusions in the context of threats to internal and external validity. For the study's replicability, we detailed the complete research method in section 3, which enables replicating every phase of this mapping study.

External Validity. Obtained results cannot be generalized to other problems domains.

5 CONCLUSIONS

In this paper, we present the results from a systematic mapping study in the field of Multiple Software Product Lines (MSPL). This paper summarizes interesting insights into the state of the art with the aims of:

- understanding what are the most covered topics of the research in MSPL;
- highlighting the open issues that need to be filled by future research;
- evaluating the level of empirical evidence reached by researcher's solutions addressing the topics,
- identifying which types of innovations are most needed today

Our classification scheme reflects the historical background of multi software product lines in the period 2010-2019 and it follows the taxonomy of the lifecycle steps of a product line. Our study highlights how, currently, researchers mainly aim at developing innovative contributions in terms of methods, models, and tools. Anyway, as the MSPL field is still in a young stage, most of these papers are only a description of approaches that are most validated in a testing scenario. Their efficacy is not still based on empirical evidence. There are no papers, in fact, that clearly describe a real validation through detailed experimental design, data collection process and validity threats analysis. Therefore, according to our study, community research is called to bridge the gap of the evidence-based research, to increase the investigation of maintenance and V&V of MSPLs and, finally, define standard models and metrics to evaluate the efficacy of the MSPLs.

REFERENCES

Bagheri, E., Ensan, F., Gasevic, D., Boskovic, M., et al. (2011). Modular feature models: Representation and configuration. *Journal of Research and Practice in Information Technology*, 43(2):109.

Clements, P. and Northrop, L. (2002). *Software product lines*. Addison-Wesley Boston.

Damiani, F., Schaefer, I., and Winkelmann, T. (2014). Delta-oriented multi software product lines. In *Proceedings of the 18th International Software Product Line Conference-Volume 1*, pages 232–236.

Dhungana, D., Seichter, D., Botterweck, G., Rabiser, R., Grunbacher, P., Benavides, D., and Galindo, J. A. (2011). Configuration of multi product lines by bridging heterogeneous variability modeling approaches. In *2011 15th International Software Product Line Conference*, pages 120–129. IEEE.

Hayashi, K. and Aoyama, M. (2018). A multiple product line development method based on variability structure analysis. In *Proceedings of the 22nd International Systems and Software Product Line Conference-Volume 1*, pages 160–169.

Hubaux, A., Heymans, P., Schobbens, P.-Y., Deridder, D., and Abbasi, E. K. (2013). Supporting multiple perspectives in feature-based configuration. *Software & Systems Modeling*, 12(3):641–663.

Kitchenham, B., Al-Khilidar, H., Babar, M. A., Berry, M., Cox, K., Keung, J., Kurniawati, F., Staples, M., Zhang, H., and Zhu, L. (2008). Evaluating guidelines for reporting empirical software engineering studies. *Empirical Software Engineering*, 13(1):97–121.

Kitchenham, B., Brereton, P., and Budgen, D. (2010). The educational value of mapping studies of software engineering literature. In *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering-Volume 1*, pages 589–598. ACM.

Krueger, C. W. (2006). New methods in software product line development. In *Software Product Lines, 10th International Conference, SPLC 2006, Baltimore, Maryland, USA, August 21-24, 2006, Proceedings*, pages 95–102.

Marinho, F. G., Andrade, R. M., Werner, C., Viana, W., Maia, M. E., Rocha, L. S., Teixeira, E., Ferreira Filho, J. B., Dantas, V. L., Lima, F., et al. (2013). Mobiline: A nested software product line for the domain of mobile and context-aware applications. *Science of Computer Programming*, 78(12):2381–2398.

Petersen, K., Feldt, R., Mujtaba, S., and Mattsson, M. (2008). Systematic mapping studies in software engineering. In *Ease*, volume 8, pages 68–77.

Pohl, K., Böckle, G., and van Der Linden, F. J. (2005). *Software product line engineering: foundations, principles and techniques*. Springer Science & Business Media.

Urli, S., Blay-Fornarino, M., and Collet, P. (2014). Handling complex configurations in software product lines: a tooled approach. In *Proceedings of the 18th International Software Product Line Conference-Volume 1*, pages 112–121.

Wieringa, R., Maiden, N., Mead, N., and Rolland, C. (2006). Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. *Requirements engineering*, 11(1):102–107.

CORPUS OF REFERENCES

The paper references that belong to our corpus can be found at the URL: *https://www.uniba.it/docenti/ardimento-pasquale/articolo/view* or it can be required by sending an email to the authors.