On the Relationship between Inspection and Evolution in Software Product Lines: An Exploratory Study

Iuri Santos Souza†, Raphael Pereira de Oliveira†, Gecynalda Soares da Silva Gomes‡ and Eduardo Santana de Almeida†
†Mathematics Institute, Computer Science Department
Federal University of Bahia (UFBA), Salvador, BA, Brazil
Email: {iuri,raphaeloliveira,esa}@dcc.ufba.br
‡Mathematics Institute, Statistics Department
Federal University of Bahia (UFBA), Salvador, BA, Brazil
Email: gecynalda@yahoo.com

Abstract—Manage the evolution in Software Product Lines (SPL) can bring some benefits such as keep the traceability between assets in core assets and products, avoid some irregular growth or decrease before it becomes a threat to the system, and also use the products feedback to improve the core asset quality. In order to understand the evolution in SPL, this paper presents an empirical study to investigate evidence between information from features non-conformities and data from corrective maintenance, based on an SPL industrial project in the medical domain. The investigation aims at tracking the features non-conformities and their likely root causes using results from two preliminary studies. The first one captured and classified the features non-conformities from features specification of nine sub-domains and the second one investigated the evolution of SPL assets along the sub-domains development. The study sample was analyzed using statistical techniques, such as Spearman correlation rank and Poisson regression models. The findings indicated that there is significant positive correlation between feature non-conformities and corrective maintenance. Sub-domains with a high number of feature non-conformities had a higher number of corrective maintenance. Moreover, sub-domains qualified as high risk have also positive correlation with corrective maintenance. This correlation allows the building of predictive models to estimate corrective maintenance based on the risk sub-domain attribute values.

Keywords-Empirical Software Engineering; Software Product Lines; Software Evolution; Software Inspection; Feature Non-Conformity; Software Corrective Maintenance

I. INTRODUCTION

Nowadays, more and more companies working on specific domains have used Software Product Lines (SPL) ideas to obtain benefits related to time-to-market, cost reduction and quality improvements [1]. In the SPL scenario, quality assurance has crucial importance [2], [3] since the assets developed can be reused in several products. However, the literature [2], [3], [4] has showed that this aspect is being poorly considered, mainly considering inspections, an important technique for quality assurance.

Software inspection is a static quality assurance technique, which can be performed on each software artifact created in the software development life-cycle (e.g. requirements, design, test cases, code, and so on) [5]. Software inspections researchers advocate that inspections can lead to the detection of anywhere between 50% and 90% of defects [6], [7]. Moreover, inspections bring also important education and social benefits [8] and provide useful data to identify root causes of problems contributing to increase the assets reliability and quality level.

On the other hand, in order to achieve the SPL benefits, the SPL adoption needs a special attention in its evolution, since the evolution of a product line and its assets is driven by maintenance within an individual product, maintenance targeted to the entire product line, and repositioning of architectural components from individual product to the product line [9], [10], [11]. These maintenance are originated from a number of different sources, such as the customers using the products, future needs predicted by the company, bug fixes and the introduction of new products in the SPL.

All of these aspects make SPL evolution management much more complex and challenging. Thus, the evolution of each core asset and product should be well managed.

Nevertheless, managing the evolution in SPL can bring also some benefits such as keep the traceability between artifacts in core assets and products [10], avoid some irregular growth or decrease before it becomes a threat to the system and also use the products feedback to improve the core asset. Thus, the management of the SPL evolution combined with other techniques, such as inspections, can be considered essential activities to the success of the line.

Although some authors [6], [7], [12] mention the importance of software inspections in every phase of software development and software evolution [13], [10], [11] in the SPL, few studies establish the relationship among inspection and evolution in the SPL context.

Thus, this paper presents an exploratory study to investigate evidence between features non-conformities information and corrective maintenance data from an SPL industrial project in the medical domain. We believe that the findings...
discusses future directions.

The remainder of this paper is organized as follows: Section II presents the related work. Section III details the context, the design and research questions of the empirical study carried out in this work. Section IV presents the results of this work. Section V discusses the main findings of the study. Section VI presents the threats to validity of the study and finally, Section VII describes the conclusions and discusses future directions.

II. RELATED WORK

In this section, we describe three studies related to our proposal, in terms of applying, understanding and evaluating the relationship between software quality and software evolution.

Marvin and Iona [14] presented an overview of the NASA Space Shuttle software Independent Verification and Validation (IV&V) process and an analysis of the Issues Tracking Reports (ITRs) produced during the IV&V process. A total of 22 versions of a large complex software system that pilots the Space Shuttle were collected describing a software product line development. The main contribution of this study was to describe a successfully implemented model for IV&V. However, as stated by the authors, the IV&V process can be too complex or too expensive to deploy depending on the organization available resources. On the other hand, our work investigates the relationship among inspection and evolution based on more common sources of data, which do not demand considerable resources from the company. These sources of data are a bug tracking system and generated documents from inspections on features artifacts.

Zhang [15] presented an study in order to explore the relationship between software metrics and software quality. The work investigated the relationship between Lines of Code (LOC) and defects through two case studies using two public defect datasets: the Eclipse dataset and the NASA dataset. Based on the results of this study, the author argues that it is possible to use defect density values, calculated from a small percentage of the largest modules, to collaborate in the prediction of the number of defects. According to the results, using typical classification techniques, it is possible to predict defective components based on LOC. Moreover, it was identified that LOC can be an useful indicator of software quality through of the defect prediction models.

Another study analyzed the distribution of bugs system from a statistical perspective using as case studies five versions of Eclipse to show how three alternative statistical distributions may fit the bug distribution [16]. The paper discusses the relative significance of using statistical distributions to model how bugs are distributed among modules. The study presents some key distributions used in the literature to model software properties, discusses about generative models to produce such data, and applies these models to exploit the mechanism of bug introduction in software modules. The authors argue that some of these alternative distributions provide both a superior fit to empirical data and a theoretical motivation to be used for modeling the bug generation process.

The investigation in this paper aims at tracking the features non-conformities and their likely root causes using results from two preliminary studies performed by the researchers from this paper. The first study captured and classified the features non-conformities from features specification of nine sub-domains and, the second one investigated the evolution of SPL assets along the sub-domains development.

To the best of our knowledge, we did not find a work exploring or investigating simultaneously empirical data and evidence from inspection and evolution results in the SPL context. Thus, the main contribution of this work is to analyze software inspection and evolution data gathered from these two previous empirical studies in an SPL industrial project.

III. THE STUDY

A. Background

1) SPL project: The SPL industrial project has been conducted in partnership with a company, which develops strategic and operational solutions for hospitals, clinics, labs and private doctor offices. It has four main products, comprising a total set of 42 modules, which are responsible for specific functions in different sub-domains (e.g., financial, inventory control, nutritional control, home care, nursing and medical assistance). Market trends, technical constraints and competitiveness motivated the company to migrate their products from single-system development to an SPL approach.

2) Inspection Activity: The assets (e.g., product map and feature specifications) built on the Scoping phase [17] of the SPL project were reviewed by inspection activities, which were performed in order to assess the contents assuring the quality of the artifacts. Based on the Fagan’s inspection process [18] and inspection practices [19], [7], [20], [6], [21], [22], [23], an adapted inspection approach to SPL context was performed by six software engineers.

The inspection activity identifies and corrects non-conformities on the feature specification in SPL scoping phase. Feature specifications non-conformities are occurrences produced in feature specification document that not are in conformance with the required quality attribute and levels of the artifact. Assuming that in the SPL project the feature specification document have to be complete, consistent and unambiguous, thus the inspection activity will identify and remove incompleteness, inconsistency and ambiguities features non-conformities occurrences. For example, the features specification from the two first iterations in the project (9 sub-domains) were inspected and as result,
the scoping analysts could fix the features specification non-
conformities found on the same phase that the problems
were identified. The features non-conformities identified
from the inspection activity were classified within 9 non-
conformities types, defined by [24], which proposes a clas-
ification for non-conformities for requirement specification,
suitable to the SPL Features specification context.

3) SPL Evolution Analysis: The company has a bug track
system called Customer Interaction Center (CIC), which
was developed by the company itself. CIC allows the company
users to register requests for adaptations, enhancements,
corrections and also requests for the creation of new mod-
ules. When registering a new request, the user must fill a
field called request type. Based on this request type, the
records from CIC were grouped according to the types
of maintenance [25] (adaptive, corrective, perfective and
preventive). In this study, only corrective maintenance was
investigated. This maintenance type corresponds to system
ersors, database errors, operating system error, error message
(e.g. general fail in the system) and, system locking / freez-
ing, according to the request type from the CIC industrial
project. Moreover, this type of maintenance was used be-
cause it is closed related to the inspection non-conformities.
The total of features, features non-conformities, corrective
maintenance and Lines of Code (LOC) of the sub-domains
are shown in Table I.

B. Empirical Study Definition

1) Data Collection: For the exploratory study, we applied
archival data methods [26], [27] to collect data and informa-
tion from the study. All collected data were treated strictly
confidential, in order to assure anonymity of the company.
Archival data refers, e.g. documents from different devel-
opment phases, organizational charts, financial records, and
previously collected measurements in an organization [26].
For this study, we used archival data to collect features non-
conformities information captured in the software inspection
activity, Lines of Code (LOC) from the sub-domains studied,
and the corrective maintenance data from CIC.

2) Analysis Procedure: This phase comprises the quanti-
tative analysis of collected data. We performed quantitative
data analysis based on descriptive statistics, correlation
analysis [28], and the development of predictive models
[29]. The objective of using quantitative analysis is to sketch
conclusions, based on the amount of collected data, which
may lead us to a clearer chain of evidence [30]. Moreover,
the relevant data from documents, assets and extracted state-
ments, as well as the observations, were grouped and stored
in the study database in order to optimize the exploration
from sources of evidence in this study [30].

3) Validity Procedure: In order to reduce the threats to
validity, countermeasures were taken in the study design and
during the whole study. All the countermeasures followed
the quality criteria in terms of construct, external and,
internal validity as discussed in [30]. Construct validity
makes use of two strategies as described:

- **Longstanding involvement**: In this strategy, the re-
searchers had a long involvement with the object of
study allowing gathering tacit knowledge which pro-
vided us with the opportunity of avoiding misunder-
standings and misinterpretations [31].

- **Peer debriefing**: It recommends that the analysis
and conclusions be shared and reviewed by other re-
searchers [31]. It was achieved by conducting the analy-
sis with three researchers, performing discussion groups
in which analysis and conclusions were discussed and
supervised by a statistical researcher (third author of
this paper).

**Internal validity**: This threat was mitigated by ensuring
the company anonymity and by the free access, within the
company to the research team.

**External validity**: Although the study was applied in only
one company, our intention is to build a knowledge base to
enable future analytical generalization where the results are
extended to cases which have common characteristics.

**Reliability**: This aspect was achieved by using two tac-
tics: a detailed empirical study protocol and, a structured
study database with all relevant and raw data such as meet-
ings tapes, transcripts, documents, and outline of statistical
models.

C. Research Questions

In order to guide this investigation, seven research ques-
tions were defined:

- **RQ1.** Is there any correlation between the features non-
conformities and corrective maintenance?
- **RQ2.** Is the sub-domain with greater number of fea-
tures or feature non-conformities the responsible for
greater number of corrective maintenance?
- **RQ3.** Is there any correlation between lines of code and
amount of features non-conformities?
- **RQ4.** Is there any correlation between lines of code
and the amount of corrective maintenance?
- **RQ5.** Is there any correlation between Risk attribute
and corrective maintenance? Is it possible to make a
predictive model of corrective maintenance for sub-
domains from Risk attribute?
- **RQ6.** What is the distribution of features non-
conformity, correctives maintenance and LOC per the
SPL sub-domain category?
- **RQ7.** What is the influence of the variables, lines of
code and corrective maintenance, on the feature non-
conformities amount?

IV. Analysis and Results of the Study

In this section, the statistical analysis and the results are
grouped by research questions. The necessary data to answer
The stated research questions were organized according to the sub-domain, as shown in Table I.

### A. RQ1. Is there any correlation between the features non-conformities and corrective maintenance?

This question investigated the correlation between features non-conformities and Corrective maintenance of the studied sub-domains in this work. The main idea was to observe if, in fact, there is any correlation between numbers of the features non-conformities and the number of corrective maintenance. The corrective maintenance, in this context, represents the system problems (errors) identified by the users.

The analysis on the investigated data set identified, as a first result, that the sub-domain with greater number of non-conformities had the greater number of corrective maintenance (sub-domain B). This evidence calls the attention for this sub-domain B as a motivation to understand what are the possible reasons to justify this evidence (is it just coincidence? or, is it a logical reflection?).

Furthermore, we tested and analyzed the correlation between the variables of this question, features non-conformities and corrective maintenance. The test was performed by Spearman’s rank correlation coefficient.

Spearman’s rank correlation coefficient or Spearman $\rho$ is a nonparametric measure for evaluating the degree of correlation between two variables [28], [32]. In this question, the variables are number of corrective maintenance ($X$) and the number of features non-conformities ($Y$). The result of Spearman correlation test is represented by the Greek letter $\rho$ (rho), where $\rho$ can have value between -1 and 1. If $\rho$ value is positive, it means that there is a positive correlation and the corrective maintenance amount ($X$) tends to increase when $Y$ increases, and when $\rho = 1$, it means that there is a perfect positive correlation. If $\rho$ value is negative, it means that there is a negative correlation and the corrective maintenance amount ($X$) tends to decrease when $Y$ increases, and when $\rho = -1$, it means that there is a perfect negative correlation. On the other hand, if $\rho$ value is 0 or close to 0, it means that there is no tendency for $X$ to either increase or decrease when $Y$ increases. Then, $\rho \approx 0$ means that there is no significant correlation between the analyzed variables. The coefficients and variables values for Spearman’s rank tests can be seen in Table II.

The correlation analysis using Spearman rank correlation test for the variables corrective maintenance and feature non-conformities presented a result within significance level of 5%, as well as it had significant correlation, once that the Spearman’s rank test return $\rho \approx 0.76$ (Table II [a]). Thus, the amount of corrective maintenance tends to increase when the features non-conformities amount increases. This tendency can be observed in the Scatterplot from Figure 1.

![Scatterplot - Corrective maintenance and Feature non-conformities](image)

**Figure 1.** Scatterplot - Corrective maintenance and Feature non-conformities

### B. RQ2. Is the sub-domain with greater number of features or feature non-conformities the responsible for greater number of corrective maintenance?

Analyzing the study data set (Table I), the sub-domain C had the greater number of features (23 features), followed by the sub-domain B with a total of 22 features. Based on the data from features non-conformities and corrective maintenance, we can observe that the sub-domain B had greater number of feature non-conformities (33 non-conformities), followed by the sub-domain C that had 31 feature non-conformities.
Table II
SPEARMAN RANK CORRELATION PARAMETERS

<table>
<thead>
<tr>
<th>Correlation</th>
<th>ρ</th>
<th>S-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>[a] Corrective Maintenance and Non-Conformities</td>
<td>0.7615</td>
<td>28.6184</td>
<td>0.0171</td>
</tr>
<tr>
<td>[b] Lines of Code and Non-conformities</td>
<td>0.1924</td>
<td>96.9036</td>
<td>0.6198</td>
</tr>
<tr>
<td>[c] Lines of Code and Corrective Maintenance</td>
<td>0.55</td>
<td>54</td>
<td>0.1328</td>
</tr>
<tr>
<td>[d] High Risk and Corrective Maintenance</td>
<td>0.7559</td>
<td>20.5020</td>
<td>0.0300</td>
</tr>
<tr>
<td>[e] Relevant Risk and Corrective Maintenance</td>
<td>0.2519</td>
<td>62.8340</td>
<td>0.5472</td>
</tr>
</tbody>
</table>

On the other hand, the sub-domain B had greater number of corrective maintenance (2473 maintenances) and the sub-domains D and C had second and third greater number of corrective maintenance (903 and 775 maintenances, respectively).

The data show that the sub-domain B is the second greater sub-domain in number of features, however, it had greater number of feature non-conformities. Moreover, the sub-domain B had greater number of corrective maintenance. Checking the possible reasons for this, it can be observed that the sub-domain B was qualified with high volatility\(^1\) (Table III) and it can justify the higher values for the numbers of corrective maintenance.

Table III
SCOPE SUB-DOMAINS SUMMARY

<table>
<thead>
<tr>
<th>Sub-domains</th>
<th>Risk</th>
<th>Volatility</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>B</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>C</td>
<td>relevant</td>
<td>low</td>
</tr>
<tr>
<td>D</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>E</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>F</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>G</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>H</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>I</td>
<td>relevant</td>
<td>low</td>
</tr>
</tbody>
</table>

C. RQ3. Is there any correlation between lines of code and amount of features non-conformities?

This question aims at investigating whether there is any correlation between the amount of Lines of Code (LOC) and the amount of feature non-conformities. In order to answer this investigation, it was applied the correlation test using Spearman correlation rank. The Spearman’s rank test indicated \(\rho \approx 0.19\) with \(p\)-value = 0.62 (Table II [b]) that did not express any correlation between the variables. Moreover, the correlation test significance level revealed a value greater than 5%. Thus, the analysis could not find correlation between lines of code and the amount of features non-conformities.

D. RQ4. Is there any correlation between lines of code and the amount of corrective maintenance?

The objective of this question is to investigate whether there is any correlation between the amount of LOC and the amount of corrective maintenance. In order to answer this investigation, it was applied the correlation test using Spearman correlation rank. The Spearman’s rank test denoted a \(\rho \approx 0.55\) with \(p\)-value = 0.13 (Table II [c]). Thus, the numbers showed by Spearman correlation test express a weak correlation between the tested variables (LOC and corrective maintenance). However, the correlation test significance level (p-value) indicated a value greater than 5%. Due to the presented p-value (greater 5%), the weak correlation between the variables could not be validated and it cannot indicate that there is correlation between LOC and corrective maintenance.

E. RQ5. Is there any correlation between Risk attribute and corrective maintenance? Is it possible to make a predictive model of corrective maintenance for sub-domains from Risk attribute?

This question intends to investigate if there is correlation between the Risk sub-domain attribute and corrective maintenance in the study data set.

In the scoping phase of the SPL project, a workshop was held with the domain experts from the organization. In this workshop, the experts qualified the sub-domains according to some attributes, such as: risk, volatility and so on [33]. In order to contextualize this question analysis, the risk attribute is described next.

Risk: These are identified and analyzed to determine the negative impact that they have on the sub-domain. In the analysis, the risks are prioritized according to the staff perception about the risk severity. The impact of risks can be:

- **High**: potential problems will occur and will be difficult to manage.
- **Relevant**: any problems can occur, however, they can be managed.
- **Low**: no apparent problems.

The analysis of this question was organized in two parts. In the first part, the correlation between the Risk domain attribute (Table III) and the corrective maintenance was tested using Spearman correlation rank. In the second part, we used non-linear regression models (Poisson) to estimate values (numbers) of corrective maintenance based on the sub-domain risk attribute value, defined by the domain experts.

\(^1\)Volatility: Determines whether the sub-domain change over time [33].
The sample used for this investigation question is composed by the CIC corrective maintenance from 2003 up 2011. In the first part of the investigation, the Spearman correlation rank revealed significant correlation just to high risk sub-domains with \( \rho \approx 0.76 \) (Table II [d]). It means that sub-domains qualified as high risk tend to have larger numbers of corrective maintenance. On the other hand, the sub-domains qualified as low or relevant risk did not show significant correlation with corrective maintenance, e.g., for relevant risk sub-domains, the correlation test showed a \( \rho \approx 0.25 \) (Table II [e]).

In the second part of the analysis, the Poisson regression analysis was applied to build prediction models to estimate the numbers of corrective maintenance based on the studied sample. The Poisson regression analysis [29] is a form of analysis used to model count data and contingency tables. It has a response variable that has a Poisson distribution, and assumes that the logarithm of its expected value can be modeled by a linear combination of unknown parameters.

In this question, the response variable (\( Y \)) is calculated from the intercept-estimated value and the unknown parameters, estimated values calculated from the risk sub-domain attribute values (high and relevant risks). Then, the linear combination is composed of intercept value (\( \hat{\beta}_0 \)) combined individually with the exponential function (\( e \)) of the unknown parameter, independent variable (\( \hat{\beta}_1 \)), and estimated amount of dependent variable (\( X \)), as can be seen in Equation 1. In this question, the estimated amount of dependent variable is one new sub-domain.

\[
\hat{Y} = \hat{\beta}_0 \cdot e^{\hat{\beta}_1 \cdot X} \tag{1}
\]

As shown in Table IV, the Poisson regression parameters values are shown: \( \hat{\beta}_0 \approx 6.22 \), and \( \hat{\beta}_1 \approx 0.65 \) for sub-domains with risk relevant and 1.47 for sub-domains with risk high. Based on the Equation 1 and calculating these values for one new relevant or high-risk sub-domain (Equations 2 and 3), it achieves the result estimated corrective maintenance values of \( \hat{Y} \approx 11.92 \) for sub-domain qualified as relevant risk and \( \hat{Y} \approx 27.13 \) for sub-domain qualified as high risk. Thus, when one new sub-domain is qualified as relevant risk, it is possible to estimate the addition of 11.92 correctives maintenance just due the fact that the sub-domain has a relevant risk. On the other hand, if the new sub-domain is qualified as high risk, it is possible to estimate the addition of 27.13 correctives maintenance due the fact that the sub-domain has a high risk.

\[
\hat{Y} = 2.01 \cdot e^{1.16 \cdot 1} \tag{2}
\]
\[
\hat{Y} = 2.01 \cdot e^{1.26 \cdot 1} \tag{3}
\]

F. RQ6. What is the distribution of features non-conformities, correctives maintenance and LOC per the SPL sub-domain category?

In order to answer this question, samples for each variable, according to Table I, are shown in box plots and then analyzed. To compare the two samples from each box plot (Non-Conformities, Corrective Maintenance and LOC), the medians, the interquartile ranges (the box lengths), the overall spreads (distances between adjacent values) and the skewness were analyzed to generalize some possible conclusions.

According to the non-conformities box plot (Figure 2), the medians are well separated, with the median for commonalities being higher, reaching 20 non-conformities. The length of the boxes is quite similar. The overall spreads are roughly similar for the two data sets. The box plot for variability shows a slight upper-skew: the upper whisker is longer than the lower. The main body of data for the commonality looks symmetric. The median for variability is lower than the low quartile of the commonality, which indicates that the number of non-conformities varies according to the asset type (common and variable). For the showed samples, common assets have more feature non-conformities than variable assets.

For the corrective maintenance box plot (Figure 3), the medians are more separated, with the median for commonalities being higher, reaching around 900 correctives maintenance. The length of the commonality box is bigger than the variability box one. The overall spreads are different, being smaller for variability. The box plot for variability shows a lower-skew: the lower whisker is longer than the upper. On the other hand, the box plot for commonality shows a upper-skew: the upper whisker is longer than the lower. The median for the variability is close to the lower adjacent value for the commonality, which indicates that the number of corrective maintenance varies according to the asset type. According to this box plot, the number of corrective maintenance for

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimate</th>
<th>Std Error</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>6.2156</td>
<td>0.0224</td>
<td>278.1094</td>
<td>&lt;2.22e-16</td>
</tr>
<tr>
<td>High Risk Sub-domains</td>
<td>1.4735</td>
<td>0.0270</td>
<td>54.5984</td>
<td>&lt;2.22e-16</td>
</tr>
<tr>
<td>Relevant Risk Sub-domains</td>
<td>0.6513</td>
<td>0.0319</td>
<td>20.3905</td>
<td>&lt;2.22e-16</td>
</tr>
<tr>
<td>AIC</td>
<td>1718.4547</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Akaike information criterion (AIC) is a measure of the relative goodness of fit of a statistical model. Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Table V

NON-CONFORMITIES, CORRECTIVE MAINTENANCE AND LOC BY COMMONALITIES AND VARIABILITIES SUB-DOMAINS

<table>
<thead>
<tr>
<th>Sub-domains</th>
<th>Number of Non-conformities</th>
<th>Number of Corrective Maintenance</th>
<th>LOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commonality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>4</td>
<td>683</td>
<td>203523</td>
</tr>
<tr>
<td>B</td>
<td>33</td>
<td>2473</td>
<td>325614</td>
</tr>
<tr>
<td>C</td>
<td>20</td>
<td>903</td>
<td>336483</td>
</tr>
<tr>
<td>D</td>
<td>31</td>
<td>775</td>
<td>331297</td>
</tr>
<tr>
<td>E</td>
<td>8</td>
<td>651</td>
<td>281276</td>
</tr>
<tr>
<td>F</td>
<td>13</td>
<td>105</td>
<td>185793</td>
</tr>
<tr>
<td>G</td>
<td>5</td>
<td>602</td>
<td>335819</td>
</tr>
<tr>
<td>H</td>
<td>7</td>
<td>143</td>
<td>190927</td>
</tr>
<tr>
<td>I</td>
<td>18</td>
<td>725</td>
<td>217902</td>
</tr>
</tbody>
</table>

| Variability |                           |                                  |     |
|-------------|---------------------------|                                  |     |
| A           | 500                       | 1000                             | 1500 |
| B           | 1000                      | 1500                             | 2000 |
| C           | 2000                      |                                  |     |

the common assets is bigger than the number of corrective maintenance for the variable assets.

Analyzing the lines of code box plot (Figure 4), the medians are well separated, with the median for commonalities being higher and very distanced from the variability median. The length of the boxes is different. The variability box is almost twice bigger than the commonality box. The overall spreads are roughly similar for the two data sets. The box plot for variability shows a slight upper-skew: the upper whisker is longer than the lower. On the other hand, the box plot for commonality shows a lower-skew: the lower whisker is longer than the upper. The median for variability is lower than the low quartile of the commonality, which leads to the conclusion that the number of LOC varies according to the asset type. Even though the variability box is almost twice bigger than the commonality box, the number of LOC for commonality is bigger than the number of LOC for variability in these samples.

According to the box plots (non-conformities, corrective maintenance and LOC), it can be observed that commonalities sub-domain category had a higher number of features non-conformities, number of corrective maintenance and, number of LOC comparing to variabilities sub-domain category. This observation shows that common assets are bigger in terms of LOC, non-conformities and maintenance, which can be a hard task to be performed since any modification in the core asset can affect the entire line [13].

G. RQ7. What is the influence of the variables, lines of code and corrective maintenance, on the feature non-conformities amount?

This question explores whether there is relationship among the study variables LOC, corrective maintenance and feature non-conformity to make a prediction model. In order to support this investigation, two prediction models using the Poisson regression technique were built.

The model intends to estimate the value of the dependent variable, feature non-conformity ($\hat{Y}$), when a new sub-domain is added, regarding the Lines of code - LOC ($\hat{\beta}_1$) and the corrective maintenance ($\hat{\beta}_2$), showed in the Equation
This study was conducted with the goal of investigate the relationship between features non-conformities information and corrective maintenance data from an SPL industrial project in the medical domain. The main findings of this study can be summarized as following:

a) The sub-domain B had greater number of feature non-conformities and the greater number of corrective maintenance for the analyzed dataset. Investigating possible reasons for this, we observed that the sub-domain B was qualified as high volatility (Table III), and then this can be a reason for the higher values for the numbers of corrective maintenance.

b) The variable risk sub-domain attribute presents significant positive correlation with the variable corrective maintenance. Consequently, it allows building predictive model (Poisson regression) using risk attribute for estimating corrective maintenance. This finding points out the risk sub-domain attribute as a significant predictor candidate for estimating corrective maintenance.

c) The commonality category of sub-domains had higher number of features non-conformities, number of corrective maintenance, and number of LOCs comparing to variability category. Observing Table V and analyzing the findings, it could be checked that common assets are bigger in terms of LOC and the maintenance could be a hard task to be performed since any modification in the core asset can affect the entire line.

d) The variable feature non-conformity indicated significant positive correlation with the variable corrective maintenance using Spearman’s rank correlation coefficient. Furthermore, corrective maintenance indicated significant influence on the number of feature non-conformities through the prediction model built on Research question 7 (RQ7).

e) The variables LOC and corrective maintenance indicated weak correlation and significance level (p-value) greater than 5% (p-value ≈ 13%). Moreover, LOC had not significant influence on the prediction model, using multiple Poisson regression models, to estimate number of corrective maintenance. This finding required other empirical results, replication of this study for validation, because some studies argue that LOC can be a useful indicator of software quality [15], [34].

VI. Threats To Validity

There are some threats to the validity of the study, which are briefly described, along with the mitigation strategy:

- Research Questions: The set of questions might not have properly covered all the aspects on the relation of SPL inspection and evolution. As it was considered a feasible threat, some discussions among the authors of this work and some members of the research group RiSE Labs were conducted in order to calibrate the
Table VI
MULTIPLE POISSON REGRESSION MODELS

(a) First model - Influence of Corrective maintenances and LOCs

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimate</th>
<th>Std Error</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.567</td>
<td>0.4521</td>
<td>3.465</td>
<td>5.3e-04</td>
</tr>
<tr>
<td>LOC</td>
<td>2.725e-06</td>
<td>1.73e-06</td>
<td>1.575</td>
<td>0.1153</td>
</tr>
</tbody>
</table>

AIC: 84.42

Signif. codes: 0 ‘∗∗∗’ 0.001 ‘∗∗’ 0.01 ‘∗’ 0.05 ‘.’ 0.1 ‘ ’ 1

(b) Second model - Influence of Corrective Maintenances

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimate</th>
<th>Std Error</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.2274</td>
<td>0.1395</td>
<td>15.963</td>
<td>&lt;2e-16</td>
</tr>
<tr>
<td>Corrective Maintenances</td>
<td>5.369e-04</td>
<td>1.015e-04</td>
<td>5.288</td>
<td>1.24e-07</td>
</tr>
</tbody>
</table>

AIC: 84.897

Signif. codes: 0 ‘∗∗∗’ 0.001 ‘∗∗’ 0.01 ‘∗’ 0.05 ‘.’ 0.1 ‘ ’ 1

questions.

- Features non-conformities classification: The literature does not present any non-conformity taxonomy for the features specification artifact. Thus, we set and used taxonomy for features non-conformities from the requirements non-conformities classification proposed by [24].

- Negative results: Some correlational analysis presented negative results. However, these should not be discarded immediately, and future analysis for these cases must be provided to increase the validity of the conclusion.

- Generalizability of the Findings: This experiment was performed based on data from just one company and it needs a broader evaluation in order try to generalize the results. A replication study is needed to allow researchers to conduct the same study to validate these results [26]. In this sense, a detailed empirical study protocol was defined and used.

VII. CONCLUSIONS

Software reuse is a key aspect for organizations interested in achieving improvements in productivity, quality and costs reduction. Software product lines, as a software reuse approach, have proven its benefits in different industrial environments and domains [35], [36].

Nevertheless, in order to achieve these benefits, quality techniques, such as quality assurance, should be performed on the artifacts, once each reusable asset can be used in different products. Moreover, SPL must deal with the evolution to keep the line updated according to the users needs and the changing environment. In this exploratory study, we could identify relationships between the data gathered from inspections and evolution, and we believe that both areas can work together to improve the evolution of the common, variable and product specific SPL assets.

Based on the dataset, we identified that: there is a significant correlation among corrective maintenance and feature non-conformities; the sub-domain with higher number of corrective maintenance also was responsible for higher number of feature non-conformities; the variable risk sub-domain attribute presents significant correlation with the variable corrective maintenance; the commonality category of sub-domains had higher number of features non-conformities, number of corrective maintenance and number of LOCs comparing to variability category and; corrective maintenance shows significant influence on the number of feature non-conformities. Moreover, this study showed that the attribute volatility can justify the amount of corrective maintenance and, that common assets should be well managed since common assets are bigger in terms of features non-conformities, corrective maintenance and LOC.

Furthermore, this work can be seen as an initial step towards the understanding of the relationship between inspection and evolution in the SPL context. As future work, we are planning to replicate this study in another company of the financial domain.

ACKNOWLEDGMENT

This work was partially supported by the National Institute of Science and Technology for Software Engineering (INES³), funded by CNPq and FACEPE, grants 573964/2008-4 and APQ-1037-1.03/08 and CNPq grants 305968/2010-6, 559997/2010-8, 474766/2010-1 and FAPESB.

REFERENCES


³http://www.ines.org.br


