Information and Software Technology 56 (2014) 839–849

Contents lists available at ScienceDirect

Information and Software Technology

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

A Systematic Mapping Study of Software Reliability Modeling

Felipe Febrero, Coral Calero, Mª Ángeles Moraga [*]

Instituto de Tecnologías y Sistemas de la Información, Universidad de Castilla-La Mancha, Paseo de la Universidad 4, 13071 Ciudad Real, Spain

ARTICLE INFO

Article history:
Received 11 July 2013
Received in revised form 18 March 2014
Accepted 18 March 2014
Available online 24 March 2014

Keywords:
Software Reliability
Modeling
Systematic Mapping
Taxonomy

ABSTRACT

Context: Software Reliability (SR) is a highly active and dynamic research area. Published papers have approached this topic from various and heterogeneous points of view, resulting in a rich body of literature on this topic. The counterpart to this is the considerable complexity of this body of knowledge.
Objective: The objective of this study is to obtain a panorama and a taxonomy of Software Reliability Modeling (SRM).
Method: In order to do this, a Systematic Mapping Study (SMS) which analyzes and structures the literature on Software Reliability Modeling has been carried out.
Results: A total of 972 works were obtained as a result of the Systematic Mapping Study. On the basis of the more than 500 selected primary studies found, the results obtained show an increasing diversity of work.
Conclusion: Although it was discovered that Software Reliability Growth Models (SRGM) are still the most common modeling technique, it was also found that both the modeling based on static and architectural characteristics and the models based on Artificial Intelligence and automatic learning techniques are increasingly more apparent in literature. We have also observed that most Software Reliability Modeling efforts take place in the Pacific Rim area and in academic environments. Industrial initiatives are as yet marginal, and would appear to be primarily located in the USA.

Contents

[*] Corresponding author. Tel.: +34 926295300; fax: +34 926295354.
E-mail addresses: ffebrero@eutelsat.fr (F. Febrero), Coral.Calero@uclm.es (C. Calero), MariaAngeles.Moraga@uclm.es (M.Ángeles Moraga).

http://dx.doi.org/10.1016/j.infsof.2014.03.006
0950-5849/© 2014 Elsevier B.V. All rights reserved.

1. Introduction

It should not currently be necessary to discuss the paramount importance that Software Reliability represents [1,2] in many sectors of industry and society. Software products have become extremely pervasive, and with them our dependence on software driven systems. Developing innovative and performable ways in which to build reliable systems is therefore a real need, and knowing how to assess the actual reliability level of any software product is of no less importance. If this is to be achieved then it is necessary to develop models that are able to assess what level of reliability can be delivered by the software systems, and this is the purpose of Software Product Reliability Modeling.

Reliability in Software Engineering (SE) is a particularly active area of research, and Software Reliability Modeling has been taking place since the early 1970s. Hundreds of Software Reliability models have emerged [2–4] and been published, but no consensus has been reached as to which models are the most appropriate to capture the complexity of Software Reliability. Moreover, the sheer number of models proposed has added more confusion to a particularly complex body of knowledge. This situation has led to the need to use advanced methodologies when analyzing this vast topic. Evidence-Based Software Engineering (EBSE) can provide this methodological support.

Evidence-Based Software Engineering [5] originated from medical research, has made a major contribution to the methodological evolution of the software engineering area. As developed in [6–8], Evidence-Based signifies the intention of replacing opinion with a scientific epistemology for the creation of knowledge. Evidence-based research is therefore the process of systematically reviewing, assessing and summarizing available research findings. Evidence is, in our context, the synthesis of the best quality scientific studies on a specific topic.

A Systematic Mapping Study (SMS) is a methodology proposed by the Evidence-Based paradigm that provides a means to systematically analyze a research topic [6,7] in order to provide an overview of a research area and allow us to identify the quantity and type of research and results that are available. Some SMS are already available in SE, and six of them are analyzed in [6]. It is also possible to add, for example, mapping for software testing [8] and for requirement specification [9], among others. The evidence (a model, a technique or a case study) is divulged in a publication paper, those papers which contribute to a systematic review are called primary studies, while systematic reviews and, in general, any study based on the analysis of previous research is a secondary study.

The principal objective of this paper is therefore to present the results of a Systematic Mapping Study designed and undertaken with the aim of identifying and categorizing a broad set of primary studies covering the work currently being considered by researchers as regards the various aspects of the reliability modeling of software driven systems, with a particular focus on Software Product Reliability.

This paper is organized as follows. Section 2 presents related works. A description of how to carry out a Systematic Mapping Study is presented in Section 3, while Section 4 shows the application of the Systematic Mapping Study developed to the context of software reliability, along with the principal results of our analysis. Section 5 presents the answers to our research questions, and our conclusions as regards the SMS are also discussed. Our conclusions and further works are then summarized in Section 6.

Table 1
Surveys on Software Reliability.

Title	Ref.	Year
An analysis of competing software reliability models	[10]	1978
Software reliability – status and perspectives	[11]	1982
Software reliability growth modeling: models and applications	[12]	1985
Software reliability models for critical applications	[13]	1991
Software reliability modeling survey	[14]	1996
Applying software reliability engineering in the 1990s	[15]	1998
Methods and problems of software reliability estimation	[16]	2006
Software reliability engineering: a roadmap	[17]	2007
Review of quantitative software reliability methods	[18]	2010

2. Related works

Reliability evaluation is focused on the modeling and analysis techniques for fault prediction purposes, and determines the optimal time at which to stop testing and release software, in addition to providing data with which to make tradeoffs between test time, reliability, cost, and performance. Software Reliability models are intended to capture the software properties and characteristics in a useful manner in order to support the aforementioned objectives.

Software Reliability Modeling can be divided into two main categories: prediction and assessment models, but both kinds of models are traditionally based on recording failed data and analyzing it with statistical inference. These stochastic models, which are better known as Software Reliability Growth Models (SRGMs), are the most conventional means of reliability analysis. They are used to describe the behavior as regards an application's failure during its testing and operational phase, and are black-box based, that is, the software system is considered as a whole, and only its interactions with the outside world are modeled without considering its internal structure. As an alternative to this conventional black-box analysis, architecture-based software reliability analysis considers the software's internal structure and explicitly relates application reliability to component reliabilities. The objective of this approach is the early assessment of the application's reliability, which is not possible with traditional models. More recently, approaches such as Bayesian Belief Networks or Test-Based methods have been mentioned in literature as being relevant categories for SRM. In general, an overwhelming number of models have been proposed in literature, and several surveys have therefore appeared. Some of the most relevant are shown in Table 1.

The structuring and classification of Software Reliability models is a considerable problem owing to the overwhelming number of models that have been proposed. Attempts have been made to deal with this complex situation, and a variety of classifications have been proposed, but to the best of our knowledge none have been developed using a systematic approach.

The strategy used to deal with this huge number of published works consisted of two stages. In the first we searched for a set of secondary studies from which to derive a generic framework. This search was systematic but sought only to find a sufficient set of studies for such a generic framework, and not an in-depth synthesis of the evidence. The selected studies are those shown in Table 1. In the second stage, the formal SMS of primary studies was conducted for a time span that was chosen as a trade-off between feasibility and relevance based on the outcomes of the first stage.

As a result of the analysis of the selected secondary studies it can be concluded that the taxonomy of the SR Models is an open question in which dozens of classification schemes have been proposed in the past.

In [10] the authors studied the most widely used reliability models, tracing the historical development of the various models and analyzing their advantages and disadvantages. They classified the models as data domain and time domain models. Some years later, the authors of [11] provided a survey of the various software reliability models that have been proposed since the early 1970s.

In this work, a classification scheme was created using the lifecycle phase during which the model is applicable as a primary key. The authors additionally discussed these models using the concepts of residual error size and the testing process used. Published in 1985 and focused on SRGM, the work presented in [12] thoroughly describes and then classifies the existing models by using the error detection rate, a concept that the authors introduce in the same work. In the early 1990s Pham and Pham [13] classified the SRM as being Deterministic or Probabilistic depending on the nature of the parameters used. Only two models (Halstead's and McCabe's) were classified as being deterministic, while the rest of the almost fifty reviewed were classified in one of the ten subcategories (Software Reliability Growth being one of them) of probabilistic models. In a later work, the history of SRM was reviewed [14]. In this work, the author classified Software Reliability models according to the kind of data the models uses: failures per time unit or time between failures, along with the re-use of the Musa-Okumoto classification scheme based on a selected set of attributes. In [15] a review of the progress in software reliability since 1975 was carried out. An interesting survey of models appears in [16], although this is mainly focused on SRGM.

More recently, in [17] three main approaches to Software Reliability Modeling were identified: the known data and time domain classes, and the error seeding and tagging approach. The IEEE Standard 1633-2008 [19] states that "there are the following three general classes of SR prediction models: exponential non-homogeneous Poisson process (NHPP) models, non-exponential NHPP models and Bayesian models". In [18] the SR models have been sorted into four classes; Software Reliability Growth Methods, Bayesian Belief Network (BBN) methods, Test-Based methods and other methods, a category which considers for example benchmark practices or metric-based methods.

Despite being, arguably, highly influential papers none of the above mention the application of any kind of systematic methodology. Moreover, the most recent paper was published in 2010, and it is therefore possible that new proposals have since appeared. Bearing all this in mind, we believed that it was necessary to carry out a Systematic Mapping Study (SMS) in order to obtain, among other interesting data, a more complete classification of Software Reliability Models by means of a recognized systematic methodology.

3. The systematic mapping technique

Reviewing literature in order to identify publications related to a specific topic is a common research task that is very often accomplished without methodological support in a non-systematic manner. This kind of informal review does not provide the means to avoid bias during the selection of the publications that will be analyzed [5,20]. It is also important to note that finding the relevant papers in the massive amount of work published is a major problem. It is therefore important to have mechanisms that can be used to summarize and provide an overview of an area or topic of interest [20] in a fair and unbiased manner.

EBSE (Empirical Based Software Engineering) therefore proposes [5–7] a variety of methods, the more relevant of which is the Systematic Literature Review (SLR) and is well-known by software engineering researchers. Systematic Mapping Studies are, on the contrary, still largely unknown [9] in SE, but it may, on occasions, be preferable to use SLRs depending on the nature of the analysis carried out.

The differences between SMSs and SLRs are thoroughly discussed in [21], in which we can read that both techniques share the same basic methodology, although they have different goals and scope. The principal goal of an SMS is the overview and classification of literature, while that of an SLR is to extract and aggregate the best information from the literature available. The scope of the studies is also different: mapping studies usually relate to a broad topic area whereas an SLR usually relates to a much more detailed subject, and there are consequently differences in the nature of the research questions. A standard systematic literature review is driven by a very specific research question while a mapping study, since it reviews a broader topic, will contain several rather generic research questions. A standard systematic literature review makes an attempt to aggregate the primary studies in terms of the research outcomes, whereas a mapping study does not discuss the outcomes of the primary studies and usually aims only to classify the primary research papers selected.

The objective of systematic mapping is therefore to provide an overview of a research area in order to assess the quantity and type of primary studies that exist on a topic of interest [6] with the aim of classifying the available research and identifying sub-topics in which more primary studies are needed. The outcome of a Systematic Mapping Study is a high level map, usually in the form of a set of tables and graphics, containing condensed information about a research area and visualizing the status of the field with regard to the research questions.

Upon considering the differences between a Systematic Literature Review and a Systematic Mapping Study, along with the objective of our work, we concluded that the best option would be to carry out a Systematic Mapping Study.

According to [20], a Systematic Mapping Study is conducted by: (1) planning, (2) conducting a search and (3) screening primary studies using inclusion and exclusion criteria. A systematic mapping also conducts (4) data extraction and analysis through the identification of categories and the classification of the primary studies in these categories. This eventually leads to the last step (5), which consists of building a map containing the results.

(1) **Systematic Mapping planning**. In this step, the systematic mapping plan that will be used as a basis to conduct the systematic mapping is established. The following tasks are typically carried out at this stage:

a. Defining the scope by means of the Research Questions. These questions need to be designed with regard to the objective that the systematic mapping is intended to attain.

b. Search strategy: Selection of sources. We establish which search sources (state-of-the-art academic and professional publication databases) will be used to find the primary studies. Availability, accessibility and quality criteria need to be taken into account at this moment.

c. Selection criteria. One element that is of paramount importance during the systematic mapping planning is the definition of the Inclusion Criteria (IC) and Exclusion Criteria (EC). These criteria make it possible to include primary studies that are relevant to answer the research questions and exclude studies that do not answer them. These criteria must be straightforward to apply and not require any interpretation to do so, in order to mitigate the cultural bias of each evaluator during the study selection step.

(2) **Conducting the search.** This signifies searching for relevant papers. In this step, the search for primary studies is conducted according to a previously established plan. This search is conducted by looking for all the primary studies that match the search string in the search sources. This can be carried out automatically if these sources have an efficient search engine. This is done by establishing the keywords and the search strings.

(3) **Selection of the primary studies.** In this step, the selection criteria (i.e., inclusion and exclusion criteria) are applied in order to select the relevant primary studies.

Table 2
Research Questions.

RQ	Description
1	How many, who carried them out, and what kinds of SRM activities have there been in the past 10 years?
2	What SRM research topics have been addressed over the past 10 years?
3	Which SRM models are in use?
4	Which terms are most often associated with Software Reliability Modeling?

(4) **Analysis and classification.** The reviewers read the title and abstract in the search for terms and concepts that reflect the contribution of the paper. While doing so, the reviewer also identifies the context of the research. The studies can be grouped into categories that cluster those which are most closely related.

(5) **Map building**. Once the classification scheme is in place, the relevant articles are sorted into the scheme, i.e., the actual data extraction takes place. The classification scheme evolves simultaneously to the data extraction through the addition of new categories or the merging and splitting of existing categories.

4. Systematic mapping application

Having presented the general steps of a Systematic Mapping Study, we shall now apply them to our work. Our systematic mapping aims to, as far as possible, identify, structure and classify those primary studies that propose Software Product Reliability Models or Theories. The mapping study process suggests that the papers should be screened by at least two researchers in order to avoid biased evaluations. This systematic mapping was conducted on February 2014 to cover papers published between 2003 and 2014 and involved three people: two software engineering researchers and a PhD candidate. The systematic mapping was conducted using the process presented in the aforementioned five steps, and was completed with the appropriate assessment of threats to validity.

4.1. Step 1: Systematic Mapping Planning

Research Questions: As explained previously, we are interested in general high level questions such as trends in research activity, discovering the concepts that are most frequently associated with our topic of interest, the sort of studies that are available, and so on.

These interests need to be explicitly expressed in the form of answerable questions. Since our intention is to identify the most important concepts in Software Product Reliability Models or Theories, the research questions shown in Table 2 have therefore been defined.

For example, in order to explore the research trends, emerging or abandoned approaches or the evolution of the research activity we could ask about the research activity during a particular time span. The following research questions are therefore proposed: How many, who carried them out, and what kinds of SRM activities have there been in the past 10 years? What SRM research topics have been addressed over the past 10 years? Will this provide us with an overview of the main fields of interest in addition to side topics in SRM research? Which SRM models are in use? These questions cover the most important designs and methods along with gaps and underrepresented approaches.

Table 3
Selected databases.

Database	Location
IEEE Xplore	www.ieeexplore.ieee.org
ACM Digital Library	www.portal.acm.org
ScienceDirect	www.sciencedirect.com

A final research question is: Which terms are most often associated with Software Reliability Modeling? With this, our intention is to identify the most important concepts in the topic. This question is also directly related to the methodology of this mapping study since "keywording" is intended to lead to the taxonomy by clustering the meaningful terms belonging to the same category.

4.1.1. Selection of sources

Three of the largest and most complete scientific databases were chosen as the sources of primary studies (see Table 3): IEEE Xplore, ACM Digital Library and Science-Direct, owing to their ease of accessibility and because, as is noted in [22,23], they are widely recognized as being an efficient means to conduct Systematic Reviews, and thus mappings in the context of Software Engineering. These databases are also commonly used sources when conducting systematic surveys in computing research. However, we considered not only effectiveness, but also the ability to export the results to a well-defined standard format in a straightforward manner.

4.1.2. Establishment of selection criteria

It is at this stage that the inclusion and exclusion criteria are formalized. The inclusion criteria of our systematic mapping are shown in Table 4 whereas the exclusion criteria are shown in Table 5.

4.2. Step 2: Conducting the search

The search is conducted using automated search engines. In order to make the search as unrestricted as possible, the search string is simply "Software Reliability". In practice we also built the particular settings for each search engine (see Table 6), since each of them works in a very specific manner. In particular, we attempted to minimize duplications and rejections by setting the appropriate options in each engine e.g. setting the ACM engine so as not to report IEEE published works that would also be reported in the output of its own engine, and also asking only for papers with an abstract or which had been published in the appropriate forums.

As a result of this step, we obtained a total of 972 papers: 114 from ACM digital library, 739 from IEEE Xplore and 119 from Science Direct (see Table 7). In all cases bibliographic data, including the abstract, were exported and stored in BibText format for further analysis. The analysis of the primary studies was supported with JabRef [24], an open source reference manager system that is able to manage, among other things, BibText databases in a very efficient manner.

4.3. Step 3: Selection of the primary studies

During the screening of the papers for relevant publications regarding our research questions, and bearing the above criteria in mind, we closely examined the title, abstract and keywords of each paper. In the case of papers whose abstracts did not provide sufficient information to make a decision about the content of the paper, it was necessary to read the whole paper to determine its relevance. The aforementioned inclusion and exclusion criteria were therefore applied at this stage. A paper would be accepted as a primary contributor to the study as long as it met at least

Table 4
Inclusion Criteria.

IC	Description
1	The paper directly relates to SRM
2	The paper evaluates one or more SRM related techniques
3	The paper compares two or more SRMs
4	The paper theoretically evaluates one or more features of an SRM technique
5	The paper discusses general or particular aspects of an SRM

Table 5
Exclusion Criteria.

EC	Description
1	The paper focuses on hardware reliability
2	The paper focuses on a software process rather than on a product
3	The paper focuses on ways in which to build reliable systems rather than SRMs
4	The paper focuses on system reliability testing rather than SRMs
5	The paper is not accessible
6	The abstract is not available
7	The document is not a paper, e.g. it is a conference cover, poster, etc.

one of the Inclusion Criteria. A paper would similarly be rejected when at least one of the Exclusion Criteria was fulfilled.

The selection process comprises five iterations. The first three were carried out by one reviewer, the fourth by two evaluators and the last one was conducted as a joint review by the three researches.

In the first iteration only papers that were clearly not within the scope of the research questions were excluded, and this first review was also very useful as regards developing an understanding of the topic. It was in the next iteration that the first set of excluded and included studies was obtained. It is important to highlight that the studies included were labeled with keywords denoting the reason for their selection. This labeling greatly facilitated any further analysis, particularly the classification work. In this step, undetermined papers (whose abstracts were not sufficiently clear) were labeled as such. A third iteration was carried out by the same reviewer with the aim of remedying eventual mistakes and deciding on those papers that had been considered as undetermined in the previous step. The full text in these papers was reviewed, and the other two researchers then reviewed the results of all the above analyses. The intention of this step was to reduce any bias in the results. Discrepancies regarding the inclusion and labeling of a paper were discussed and agreed on in a final joint review.

A total of 503 papers were eventually selected – 386 from IEEE Xplore, 50 from ACM Digital Library and 67 from ScienceDirect (see Table 7).

Table 6
Search string.

Database	Search string and settings
IEEE Xplore	You searched for: (“Document Title”:”software reliability” OR “Abstract”:”software reliability”) You Refined by: Publisher: IEEE Content Type: Conference Publications, Journals & Magazines Publication Year: 2003–2014
ACM digital library	(Abstract:”software reliability”) and (not Publisher:IEEE) and (PublishedAs:journal OR PublishedAs:proceeding OR PublishedAs:transaction) and (FtFlag:yes) and (AbstractFlag:yes) Publication Year: 2003–2014
Science Direct	pub-date > 2002 and TITLE-ABSTR-KEY (“Software Reliability”) [All Sources (Computer Science, Engineering)]

Table 7
Selection summary.

Source	Included	Excluded	Total
IEEE Xplore	386	353	739
ACM Dig. Lib.	50	64	114
ScienceDirect	67	52	119
Total	503	469	972

It should be noted that the number of duplicities was very low, as expected with the search settings. Moreover, it is worth mentioning that a notable set of the excluded works was sorted into two categories that can be defined as “ways to attain reliable systems” (which refers to, e.g., techniques such as rejuvenation or how to build fault-tolerant software) and “software reliability testing” e.g., testing techniques or testing process management) in addition to those that were not papers.

4.4. Step 4: Analysis and classification

As mentioned previously the SMS methodology preconizes the classification of the primary studies by applying keyword clustering. This means that the classification categories are obtained using the keyword clustering of the meaningful terms extracted from the title, abstract and author’s keywords when provided by the authors. This process thus relies on what the authors claim about their own work. The keyword clustering is carried out in two steps. Once the meaningful terms or keywords have been extracted they are clustered into relevant categories with regard to the research questions, e.g. the terms “prediction” and “assessment” are keywords that are used to classify the works as being related to the prediction or assessment of software Reliability.

A classification scheme of this nature can be developed from scratch by considering only the set of included primary studies, or constructed by relying on previously proposed taxonomies for the topic under study. We have chosen the option of driving our classification criteria with a previously proposed taxonomy since we consider it the better approach to deal with the sheer heterogeneity of terms in the context of an SMS. We chose the proposal depicted in [18] because we found that the proposed categories were easy to identify without the need to analyze each paper in depth as occurs in SLRs but also because it is a very recent proposal that has originated from a reliable institution.

The models were therefore first sorted according to [18] into four classes:

1. Software Reliability Growth Models (SRGM).
2. Bayesian Belief Networks (BBN).
3. Test-Based methods (TBM).
4. Other methods.

These were the basic categories, but our analysis, which was driven by the keywords in the title, abstract and author’s keywords that provide information about the nature of the work, was open to the addition of new categories if the keyword clustering suggested that this was pertinent. As will be seen later, this was indeed the case. Specifically, we realized that a large percentage of works were classified into the category “others”. Upon analyzing the clustering of meaningful terms, two groups containing a considerable amount of works were identified. It was therefore considered appropriate to add two new classes to the classification presented above. The taxonomy resulting from this modification will be presented in the following section together with the rest of the mapping outcomes.

It should be noted that it is not exceptional to encounter publications that present two or more issues, such as a new mathematical

Table 8
Main conferences.

Acronym	Conference name	Number of papers
ISSRE	Intl Sym on Software Reliability Engineering	28
COMPSAC	Computer Software and Applications Conference	20
IEEM	Industrial Engineering and Engineering Management	15
PRDC	Pacific Rim International Symposium on Dependable Computing	14
RAMS	Reliability and Maintainability Symposium	14
ICRMS	Intl Conf Reliability, Maintainability and Safety	12
SSIRI	Secure Software Integration and Reliability Improvement	10

technique and its application to one or several well-known models, or a new or enhanced model and its experimental application. In these cases, our classification chose what we considered to be the most relevant contribution of the various issues presented as the key for the categorization. This was done because we considered that it would be more difficult to sort these works into more than one category.

Once the classification task had been accomplished, the results obtained from the systematic mapping were analyzed in order to answer the research questions. The analyses of the results are the objective of step 5, and they will therefore be presented during the description of this step.

4.5. Step 5: Map building

The most obvious fact to emerge when looking at the final set of primary studies is that the majority of them (77%) are from conference proceedings and only 23% are from journal articles. Journal papers had principally been published in the Journal of Systems and Software (23 papers), and IEEE Transactions on Reliability (19 papers). Next are the IEEE Transactions on Software Engineering which contributes with 9 studies, and the Reliability Engineering & Systems Safety journal with 8 primary studies. The remaining papers appear in various journals but with no significant amount of instances. With regard to the main conferences, as shown in Table 8 the International Symposium on Software Reliability Engineering is the major source of the primary studies included.

With regard to the main research groups in this topic, and considering the main author as an index, we found that there were 256 different main authors in the 503 studies selected. Of these, only 18 had published more than three works during the time span analyzed.

Upon analyzing the geographical distribution it can be noted that China, followed by the USA, are the countries in which most scientific production has taken place as regards SR (see Table 9). Moreover, if we consider the contribution of the Pacific Rim area, the percentage of the total amount of works included rises of up to 76%.

With regard to the evolution in the number of papers published, we found that the amount of activity in this topic area has

Table 9
Number of papers by geographical area.

Geo area	Number of papers
China	152
USA	86
Japan	59
Taiwan	48
Europe	49
India	38
Asia (Others)	23
Mid East	11
Rest	37

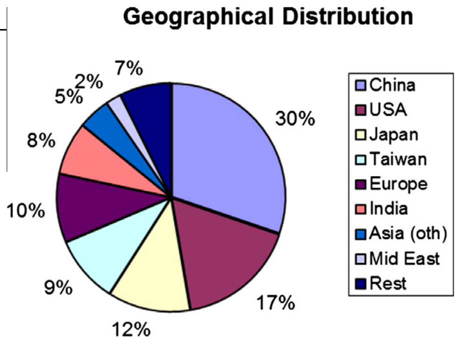


Table 10
Publication tendency by year.

Year	Primary studies
2003	28
2004	21
2005	35
2006	44
2007	53
2008	61
2009	60
2010	63
2011	65
2012	34
2013	39

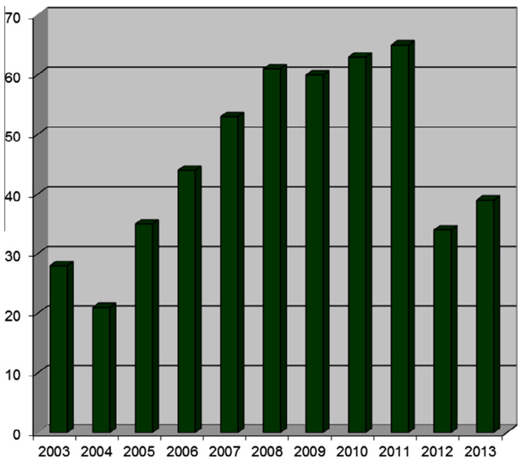
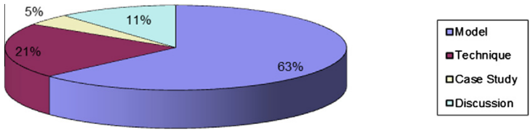


Table 11
Nature of the work.

	Model	318
	Technique	106
	Case study	22
	Discussion	57

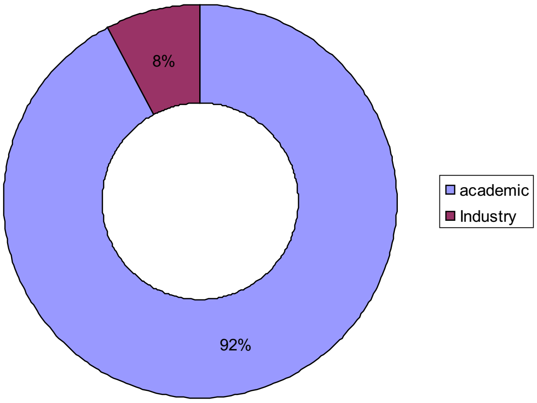


Fig. 1. Academic proposals vs. industrial proposals.

increased in recent years, particularly from 2005 on. However, since 2012 the number of papers published on this topic appears to have drastically decreased. Table 10 shows the publication frequency from 2003 until 2014.

If attention is now paid to the nature of the work, it will be noted that, as is shown in Table 11 the majority of the primary studies selected claim to present new or enhanced models.

With regard to the origin of the papers, that is, whether they have been produced in an academic or industrial environment, Fig. 1 shows that most of them appertain to an academic environment. Indeed, only 40 papers of the total of 503 papers included originated from an industrial environment, i.e., only 8% of the total.

If we focus on those works that are of our most interest to us, i.e., those 318 regarding Reliability Models, it can be highlighted that most modeling efforts are in a predictive area: 216 of the selected primary studies are focused on predicting reliability as opposed to just 102 that deal with assessment methods and models.

The various works that propose a reliability model have been initially classified according to the SRM taxonomy of [18] presented in step 4 in this section. However, during the study we found up to 141 works that did not fit into any of the three specific classes, signifying that approximately 44% of the selected papers had to be classified as "others". This option was discarded for two reasons: (1) The percentage of papers is sufficiently representative of the total to be considered as more than simply "others". (2) We have used keyword clustering to identify that a significant number of them, 89, could fit into new classes.

The way in which the SMS methodology runs the keyword clustering, or grouping of meaningful terms, is fairly simple. First we extract a list of the most meaningful terms or "keywords" and then we identify the groups or "clusters" of these keywords. For example, in the IEEE Taxonomy terms or keywords such as back-propagation, machine learning, statistical learning, greedy algorithms, support vector machines, particle swarm optimization, genetic algorithms or fuzzy logic (among others) are classified in the "Computational and Artificial Intelligence" category. In a similar way, as we noted that a significant number of works that had initially been classified as "other" contained these kinds of terms, we concluded that it would be appropriate to define a separate category for them. As a consequence of this process we decided to add two more classes to the classification provided by [18] because, from our point of view, the new taxonomy would better reflect the reality of the papers under study. These classes are those which correspond to the models inspired by AI based techniques and models built on the consideration of the Static structure and Architecture of the product, namely Static & Architectural Reliability models. We have also slightly modified the scope of the BBN class to include non-SRG models and those that are in some way related to the Bayes Theorem and bayesian inference other than the solely Bayesian Belief networks, and we consequently renamed this class as Bayesian Methods.

The following new taxonomy was therefore obtained:

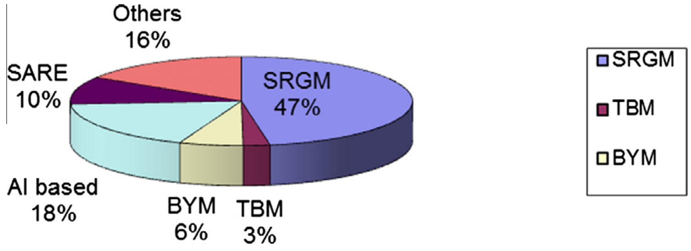
1. Software Reliability Growth Models (SRGM).
2. Bayesian Methods (BYM).
3. Test-Based methods (TBM).
4. Artificial Intelligence based techniques (AI based).
5. Static and Architectural Reliability models (SARE).
6. Other methods.

The well-known SRGMs are those time based models that, by means of an empirical approach first used in hardware reliability, describe how the underlying defects affect the observation of failures. The BYM are methods that, by means of a Bayesian Inference, account for the influence or effects of one event on another. These are often implemented as Bayesian Belief Networks and are used for modeling by combining disparate information. The Test Based Methods are specifically interested in considering the way in which the software is used. The selection of test cases should be based on the software's Operational Profile which reflects the softwarés input space. AI based methods are those inspired by Artificial Intelligence techniques, typically Neural Networks, but also other automatic learning or reasoning concepts. Finally, SARE models are built on both the consideration of the architecture and the static features of the software, and correlate software engineering measures and SR. Such models are intended to be used during the early stages of the Life Cycle, before the actual software is available. We also have a class for "other" methods and models that contains ad hoc techniques based on particular theories, such as Reliability Block Diagrams, Risk Analysis, Failure Mode Analysis, the Deterministic Chaos Theory and the Cloud Model Theory, or which were developed for very specific cases, such as an ad hoc model for a particular N-Version programming system or methods based on engineering metrics and benchmarking or early software reliability assessment, based on software behavioral requirements.

The classification of the papers based on our taxonomy is presented in Table 12.

Table 13 provides details on the percentage of proposed models by year and class. This has allowed us to obtain that there is a similar tendency in all the years studied. In general, the type of model most frequently used is SRGM, whereas those least used are TBM and BYM.

Table 12
SR models classes.



Class	Number
SRGM	150
BYM	19
AI based	57
SARE	32
TBM	8
Others	52

Table 13
Percentage of proposed models by year and class.

Year	SRGM	SARE	TBM	BYM	AI based	Others
2003	76	8	0	8	8	0
2004	78	7	7	0	0	7
2005	54	11	0	7	21	7
2006	53	12	0	9	17	9
2007	39	3	6	3	26	23
2008	42	15	0	5	15	23
2009	35	18	0	9	23	15
2010	32	2	7	9	27	23
2011	65	5	2	0	13	15
2012	50	14	4	4	14	14
2013	25	15	0	10	15	35

Fig. 2 provides a graphic representation of the evolution. As can be seen, although the classical SRGM approach is still the most representative, it would appear that the other techniques are gradually gaining on the SRGM approach, although they are less representative. It is also important to note the significant appearance of other approaches such as the Artificial Intelligence and the Static and Architecture based approaches, which have grown rapidly since they first appeared.

We have also created a Keyword cloud using the various terms and keywords that recurrently appear in association with the SRM topic, in which the most frequently used terms are written in larger letters, thus showing on which areas the works are focused (see Fig. 3).

Having obtained all the results, in the following section we shall answer the questions posed in Section 2, Table 2.

5. Discussion

This mapping study reports on the high level aspects being researched as regards the modeling of reliability in software systems. This research has been conducted in a systematic manner and with minimal bias through the formulation of a set of research questions which, having achieved the mapping, can now be answered.

5.1. RQ1: How many, who carried them out and what kinds of SRM activities have there been in the past 11 years?

Our results (Table 10) suggest that activity in the topic is steadily increasing, although the anomalous data for 2012 and 2013 need further analysis to assess whether a real change in trends has occurred. We do not believe that this is the case, but rather that other reasons such the replacement of key terms with other emerging terms might better explain this anomaly, i.e., the term ‘dependability’ is now sometimes used with a similar sense to ‘reliability’.

As indicated previously in Fig. 4a, the initiatives on SR modeling are largely led by academic teams principally located in the Pacific Rim area, with an important prevalence of China, Japan and Taiwan in addition to the notable contribution of the USA. It is also interesting to point out that the majority of industrial proposals, Fig. 4b, have been proposed in the USA and Europe. It can therefore be highlighted that there are differences between the countries that lead research depending on the environment.

We should also note the high dispersion of working groups, 256 for 503 papers, in addition to the fact that only 18 of the 256 main authors had signed more than three works, which has led us to consider that many of the works published have no real continuity and that their development ends just after being published. The low concentration of publishing forums (note that the main contributions in either journals or conferences consist of barely twenty papers), together with their heterogeneity is also noteworthy. All of the above can be interpreted as signs of the discipline’s immaturity.

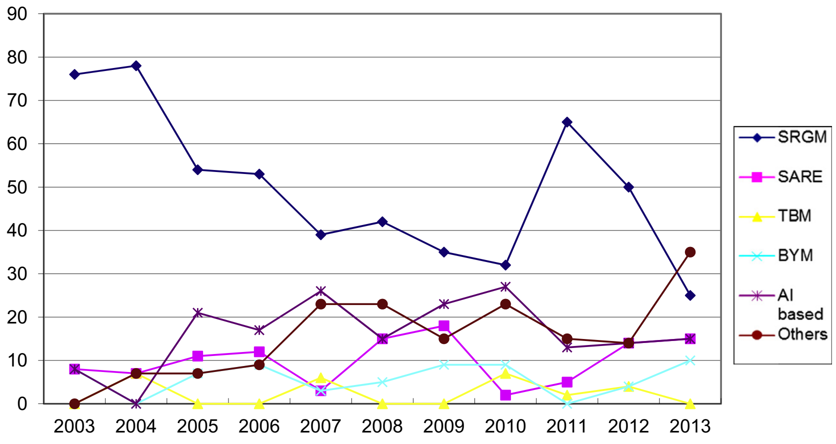


Fig. 2. Evolution of Reliability Models studies.

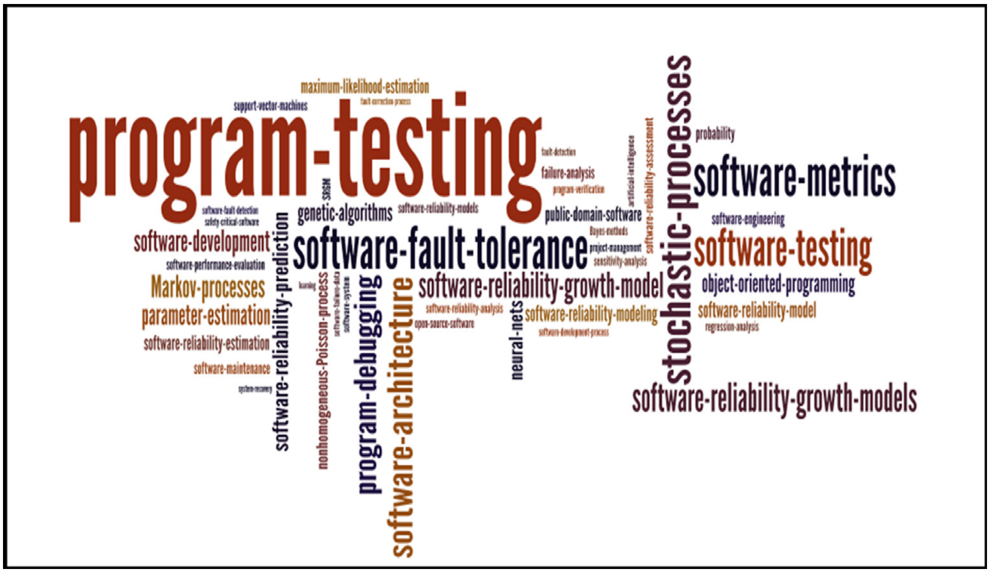
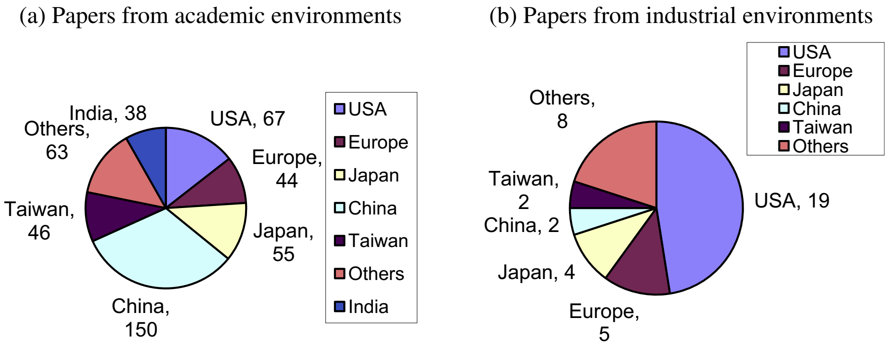


Fig. 3. Weighted topic cloud, created with http://www.wordle.net/.


(a) Papers from academic environments

(b) Papers from industrial environments

Fig. 4. Geographical distribution of papers.

5.2. RQ2: What SRM research topics have been addressed over the past 11 years?

The first conclusion is the still important prevalence of SRGM (Table 12). It is also important to note that 65% of the works studied (Table 11) deal with either (1) a proposal for a new model or an enhancement of previous ones or (2) the application of one or several of them. What is more, most of the works still focus on the classical black-box statistical approach. It is both important and interesting to point out that this study shows that Artificial Intelligence techniques and static architecture based models are perhaps emerging as the next step in this topic's evolution. Another relevant piece of data is that the third class related to number of studies is "others", which reflects enormous heterogeneity, and is perhaps a sign of a lack of maturity, not as regards the topic itself but as regards the definitions of its generalized and accepted foundations. Please recall that this class typically accommodates adhoc models that are only applicable in particular circumstances, or models based on very particular techniques. It is also worth highlighting that the authors were surprised to find so few modeling activities based on normative proposals such as international standards like the ISO/IEEC 25010 [25]. It is of great importance to carry out further research on this in order to avoid the problem of a lack of consensus on this topic's foundations.

5.3. RQ3: Which models are in use?

Our study revealed (Table 11) that very few real-world Case Studies (CS) have been published. The number of CS is marginal when compared to the total number of papers presented. Although it is increasingly common to find that theoretical proposals are completed with application examples, it is often the case that they cannot be considered as real CS. We have also observed that only the 8% of the total number of papers included originated in industrial initiatives (see Fig. 1).

Upon analyzing these works we noted that many of them are empirical ad hoc modelings or lesson learning reports on very specific reliability analyses. We further noted that SRGM are the most frequently applied or analyzed in industrial environments (roughly a third). There are three examples of industrial initiatives that apply BBN, and two ANN applications. The remaining works (the last third) are either discussions on a variety of reliability related subjects, e.g. data quality and the customers' viewpoint, or are the automated generation of test cases which are proposals that are based on, for example, exotic theories such as the Chaos Theory, Risk analysis or SFTA, in addition to several comparative studies.

From all of the above it would appear clear that a barrier preventing the generalized use of the SRM proposal in industry still exists.

5.4. RQ4: Which terms are most often associated with Software Reliability Modeling?

As mentioned previously, the diversity of terms is important and represents an obstacle to the achievement of the maturity of this topic. This phenomenon is illustrated in the weighted cloud presented in Fig. 3, in which it is possible to observe the density of terms, although those most frequently used are: program testing, software fault tolerance, software metrics and stochastic processes.

5.5. Threats to validity

Despite the fact that this study has been carried out by following the SMS methodology, there may be some threats to its validity. The principal limitations of this study that have been identified concern the limited access to sources, a circumstance that may have led to a bias in the selection of publications owing to the possible existence of interesting studies in other databases. However, the databases used cover the area of software engineering well [21,22] and we have no reason to believe that this does not apply to software reliability. We are, therefore, reasonably confident that we are unlikely to have missed many significant published studies.

Another threat lies in the difficulties involved in selecting and classifying the primary studies selected owing to the lack of an appropriate taxonomy and the wide diversity of terms used in the topic of interest. This may also have led us to miss some relevant studies since the search string consisted of the term ‘‘Software Reliability’’. However, since this is, on the whole, the most commonly used term, the possibility of any significant contribution not mentioning the words ‘‘Software Reliability’’ in the title, abstract or keywords is minimal. With regard to the classification, we have added new categories to the existing taxonomy, and we therefore believe that the classification of the papers is mostly accurate.

As previously mentioned, only one author coded all the papers, and some bias could therefore have been introduced in the results for a variety of reasons such as the author’s subjectivity. However, the other authors’ supervision as regards reviewing both the included and the excluded papers leads us to believe that this error, if it exists, is minimal.

Some relevant papers might not have been found in the digital databases when using our search and selection protocol. Automated searches rely on both search engine quality and how researchers write their abstracts. Although we are reasonably confident as regards how well digital databases classify and search indexed work, if abstracts and keywords are of poor quality it is clear that the search will be greatly flawed.

Finally, only studies published in the English language were selected in the search, signifying that potentially important studies published in other languages have not been considered. However, since English is the most widely adopted language as regards writing scientific papers, the eventual bias owing to this issue is minimal.

6. Conclusions and future work

The main contribution of this work is to provide a Systematic Mapping of Reliability Modeling in software systems. This has been done methodologically, by employing a currently standard technique for Systematic Mapping Reviews in order to warrant the quality of the analysis.

A total of 972 works were obtained, 503 of which were considered as relevant publications when taking into account our research questions. These works were classified according to publication attributes, the nature of the work, topic and content features. The principal results attained have allowed us to observe that there are many research groups publishing very different initiatives for SR modeling using a variety of approaches and points of views. The lack of consensus on this issue is well-know, and it does not appear to have evolved towards an increased consensus over the last decade. We have attempted to use a classification proposed in literature but have found that it was not sufficient to classify the papers selected, and we have therefore proposed the following new taxonomy:

1. Software Reliability Growth Models (SRGM).
2. Bayesian Methods (BYM).
3. Test-Based methods (TBM).
4. Artificial Intelligence based techniques (AI based).
5. Static and Architectural Reliability models (SARE).
6. Other methods.

Surprisingly, the two new classes were the second and the third as regards those in which most proposals were found (after the classic SGRM approach). We can therefore conclude from our study that new lines of research based on both the AI domain and the static structure of the software are producing a notable amount of new evidence. Besides extending the reference classification scheme to accommodate the class of (in a broad sense) Artificial Intelligence models and the class of Static and Architectural Reliability models, the paper has also mapped trends in published research by year and by clustering the topic’s terminology.

Another important point that should be noted is that few real-world case studies have been published, and most model application works have, to date, focused on academic experiments. More research is therefore required in order to understand the problems of applying SRM techniques in real-world contexts.

Our mapping study has also discovered a possible gap in research since no significant work appears to have been carried out as regards modeling reliability by following the proposals in international standards such as ISO/IEC 9126 or the ISO/IEC 25K family [26].

Our future work will be to extend the study by using a systematic literature review involving more specific topics, as our position is that the issue of the applicability of the models to real-life environments needs to be addressed. As we also believe that standard-inspired reliability modeling will play an important role in achieving the ‘‘real life’’ goal, this SLR will seek software reliability models based on standards. The final goal will be to propose a Reliability model that is, as far as possible, based on standards and which will then be applied to real systems.

Acknowledgements

This research has been funded by the following projects: GEODAS-BC project (Ministerio de Economía y Competitividad and FondoEuropeo de Desarrollo Regional FEDER, TIN2012-37493-C03-01).

References

[1] B. Littlewood, L. Strigini, Software reliability and dependability: a roadmap, in: A. Finkelstein (Ed.), The Future of Software Engineering, ACM Press, 2000, pp. 177–188.
[2] M.R. Lyu, Handbook of Software Reliability Engineering, IEEE Computer Society Press and McGraw-Hill, 1996.
[3] B. Littlewood, Theories of software reliability: how good are they and how can they be improved?, IEEE Trans Softw. Eng. SE-6 (5) (1980) 489–500.
[4] J.D. Musa, D. John, K. Okumoto, Software reliability models: concepts, classification, comparisons, and practice, in: Proceedings of the Electronic Systems Effectiveness and Life Cycle Costing Conference, 1982, pp. 395–424.
[5] T. Dyba, B. Kitchenham, M. Jorgensen, Evidence-based software engineering for practitioners, IEEE Softw. 22 (1) (2005) 58–65.

[6] D. Budgen, M. Turner, P. Brereton, B. Kitchenham, Using mapping studies in software engineering, Proc. PPIG 2008 (2008) 195–204.

[7] K. Petersen, R. Feldt, S. Mujtaba, M. Mattsson, Systematic mapping studies in software engineering, in: Proceedings of the 12th International Conference on Evaluation and Assessment in, Software Engineering, 2008, pp. 68–77.

[8] W. Afzal, R. Torkar, R. Feldt, A systematic mapping study on non-functional search-based software testing, in: Proceedings of 20th International Conference on Software Engineering and Knowledge, Engineering (SEKE), 2008, pp. 1–6.

[9] N. Condori-Fernandez, M. Daneva, K. Sikkel, R. Wieringa, O. DiesteTubío, O. Pastor, A systematic mapping study on empirical evaluation of software requirements specifications techniques, in: Proceedings of Third International Symposium on Empirical Software Engineering and Measurement, 2009, pp. 502–505.

[10] G.J. Schick, R.W. Wolvertom, An Analysis of Competing Software Reliability Models, IEEE Trans. Softw. Eng. SE-4 (2) (1978) 104–120.

[11] C.V. Ramamoorthy, F.B. Bastani, Software reliability-status and perspectives, IEEE Trans. Softw. Eng. 8 (2) (1982) 354–371.

[12] S. Yamada, S. Osaki, Software reliability growth modeling: models and applications, IEEE Trans. Softw. Eng. 11 (12) (1985) 1431–1437.

[13] H. Pham, M. Pham, Software reliability models for critical applications, Report EG&G-2663.Idaho National Engineering Laboratory, 1991.

[14] W. Farr, Software reliability modeling survey, in: Handbook of Software Reliability Engineering, USA, 1996, pp. 71–117.

[15] W. Everett, S. Keene, A. Nikora, Applying software reliability engineering in the 1990s, IEEE Trans. Reliab. 47 (3-SP) (1998) 372–378.

[16] I. Karanta, Methods and problems of software reliability estimation, VTT Working Papers 63, VTT Technical Research Centre of Finland, 2006.

[17] M.R. Lyu, Software reliability engineering: a roadmap, in: Proceedings of Future of Software Engineering (FOSE'07), 2007, pp. 153–170.

[18] T. Chu, M. Yue, G. Martinez-Guridi, J. Lehner, Review of Quantitative Software Reliability Methods, Brookhaven National Laboratory Letter Report JCN N-6725, 2010.

[19] IEEE STD 1633, 2008, IEEE Recommended Practice on Software Reliability.

[20] B. Kitchenham, S. Charters, Guidelines for Performing Systematic Literature Reviews in Software Engineering, Technical Report EBSE 2007–001, Keele University and Durham University Joint, Report, 2007.

[21] B. Kitchenham, D. Budgen, P. Brereton, Using mapping studies as the basis for further research – a participant observer case studies, Inform. Softw. Technol. 53 (2011) 638–651.

[22] P. Brereton, B. Kitchenham, D. Budgen, M. Turner, M. Khalil, Lessons from applying the systematic literature review process within the software engineering domain, J. Syst. Softw. 80 (4) (2007) 571–583.

[23] T. Dyba, T. Dingsoyr, G.K. Hanssen, Applying systematic reviews to diverse study types: an experience report, First International Symposium on Empirical Software Engineering and Measurement, 2007, pp. 225–234.

[24] http://jabref.sourceforge.net/

[25] ISO/IEC 25010, Systems and software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Software product quality and system quality in use models, ISO, 2010.

[26] ISO/IEC 25000, Systems and Software Engineering – Software Product Quality Requirements and Evaluation SQuaRE.