

An Experience-based Framework for Evaluating Alignment of Software Quality Goals

Panagiota Chatzipetrou^a, Lefteris Angelis^{a,b}, Sebastian Barney^{c,d}, Claes Wohlin^c

^a Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece

pchatzip@csd.auth.gr, lef@csd.auth.gr

^b Corresponding author: tel/fax +302310998230

^c Blekinge Institute of Technology, Karlskrona, Sweden

sebastian@unswalumni.com, claes.wohlin@bth.se

^d School of Information Systems Technology and Management, University of New South Wales, Australia

Abstract

Efficient quality management of software projects requires knowledge of how various groups of stakeholders involved in software development prioritize the product and project goals. Agreements or disagreements among members of a team may originate from inherent groupings, depending on various professional or other characteristics. These agreements are not easily detected by conventional practices (discussions, meetings etc.) since the natural language expressions are often obscuring, subjective and prone to misunderstandings. It is therefore essential to have objective tools that can measure the alignment among the members of a team. Especially critical for the software development is the degree of alignment with respect to the prioritization goals of the software product. The paper proposes an experience-based framework of statistical and graphical techniques for systematic study of prioritization alignment, like hierarchical cluster analysis, analysis of cluster composition, correlation analysis and closest agreement directed graph. This framework can provide a thorough and global picture of a team's prioritization perspective and can potentially aid managerial decisions regarding team composition and leadership. The framework is applied and illustrated in a study related to global software development where 65 individuals in different roles, geographic locations and professional relationships with a company, prioritize 24 goals from individual perception of the actual situation and for an ideal situation.

Keywords: software; quality; global software development; cluster analysis; alignment; different perspectives

1. INTRODUCTION

The various groups of stakeholders involved in the software development process in an organization may prioritize quality goals differently. The stakeholders represent various roles, some of them are internal to the organization (insourcing) while others are in outsourcing relationships, and moreover, they may work remotely in different countries, representing onshore and offshore relationships. For value-based software engineering, the need for a common understanding of the prioritization of goals by key stakeholders is crucial.

The conventional methods of communication and interaction such as discussions and meetings, even with the aid of technology (online communication, email, etc), do not always guarantee that all the individuals working on the same project are aligned in their views and goals. It is therefore quite possible that managers or group leaders may be not so well informed of the real views of the members of one or more teams working on the same project. The numerous differences among stakeholders, especially the cultural and language diversities and the personal conflicts may be a source of poor communication, misunderstandings, arguments, negative climate within teams and eventually the cause of undesirable results such as unpredictable delays and poor product quality.

The aforementioned considerations, which were often observed in practice in software development professional environment, motivated the present paper. Specifically, it is our belief that it is essential to develop a statistical framework that, based on the experience, can systematically study and determine the level of alignment between success-critical stakeholders in terms of the priorities given to goals. The scope of such a framework is to aid decision makers to obtain a global picture, a kind of “*map*”, of their teams’ perspectives. Based on this valuable knowledge, which of course should be combined with the decision makers’ own experience and personal knowledge of the project’s peculiarities, decision makers can detect and solve various problems caused by the lack of common goals.

Therefore, the purpose of this paper is to introduce a framework of appropriate statistical and graphical methods for the systematic study of prioritization alignment among stakeholders belonging to inherently different groups. In this regard, Global Software Development (GSD) is a very challenging field to apply and illustrate such a framework, since it can provide data rich in information regarding several different types of stakeholders, their locations, roles and relationships with the company. Furthermore, the scale of products and the complexity of development in the GSD context involve several goals in different levels.

GSD (Šmite et al., 2010) is software work undertaken at geographically separated locations across national boundaries in a coordinated fashion involving real time (synchronous) and asynchronous interaction. While teams are not co-located, they are still working towards a common goal with a commercially viable product (Conchúir et al., 2009). GSD usually comes with a number of benefits and

risks. It is a common practice, which allows companies to improve the product management and the quality of their products. From a practical perspective several interesting questions emerge: Do the stakeholders form groups of individuals with similar views? How many groups are formed in a specific project? What is the content of each formed group? Are these groups related to other characteristics of individuals? Are there any isolated individuals, in the sense that their views are unique and divergent from the rest of the team? Or, on the contrary, are there any individuals who are “hubs”, in the sense that several others are aligned with their view? All of these questions cannot be answered by conventional methods. It is unfeasible to create a type of “*global picture*” of all individuals’ views by discussions and meetings. On the contrary, systematic study of data obtained by appropriate tools (in our case a voting scheme) can build such a global picture. Based on that, the ability to understand the levels of alignment between stakeholders can assist software development organizations to create working environments where priorities with respect to quality are well defined and understood by everyone. The research questions that we address in this paper and which are relevant to the abovementioned practical questions will be described in detail in Section 3. These research questions concern the ability of the statistical framework to address: (a) the finding of stakeholders’ groupings according to their prioritization agreement; (b) the different composition of groups; (c) the differences of goals between the groups; (d) the consistency of prioritization under different situations and (e) the graphical illustration of each stakeholder’s closest aligned person.

The methods proposed with the statistical framework first investigate how different people involved in the software development process can be clustered, based on their prioritization of goals. Second, the compositions of clusters are studied with respect to the inherent groupings of the individuals. Third, correlation between the prioritization under the different situations, actual and ideal, is explored separately for each group, where ideal represents the perceived optimal set of priorities for the actual situation. The actual situation is the currently perceived situation by individuals. The ideal is used to both identify similarities and differences of how people want the priorities of the goals to be, which helps the organization to identify different opinions of the prioritization. Finally, the notion of closest alignment is introduced and is modeled through directed graph visualization. For all the aforementioned purposes, we use data from an industrial survey from one company and propose statistical methodologies and graphical tools which are based on metrics of distance.

The empirical study uses data from a prioritization process, where 65 stakeholders in different roles and groups, working on the development of the same product, prioritize with cumulative voting 24 goals from their individual perception of two situations: the actual situation and their perceived ideal situation.

Hierarchical cluster analysis with dendrograms is applied to the prioritization data, separately for each situation and the composition of the clusters, regarding the distribution of the inherent groupings within

each of them, is analyzed. Consistency of clusters under the two situations and associations between inherent groups and formed clusters are tested using the chi-square test. Differences in prioritization of goals between formed clusters are explored with box-plots and tested with the Mann-Whitney test. Correlation between the two perspectives regarding prioritization of goals is analyzed using Spearman rank correlation coefficient. Finally, a "closest alignment" relation among stakeholders is presented, showing each stakeholder as a node in a graph, connected to the best aligned individual among the other stakeholders.

The stakeholders who participated in the study had various roles and different relationships with the company. However, not all of the combinations were present. This is because the data concern teams working on a real software product and the whole setup is not an experimental design. Specifically, there were onshore-insourcing, onshore-outsourcing and offshore-insourcing arrangements. There was no stakeholder with offshore-outsourcing relationship. The statistical and graphical methodologies applied to the data gave interesting results on the participants' views.

In brief, the statistical framework provides support in identifying clusters as well as gives other interesting input to decision makers. In this particular illustrative case, the cluster analysis resulted, for both situations, in two clusters. Although some participants from the same role tend to cluster together (e.g. Testers and Architects) in general the compositions of the resulting clusters do not differ significantly. The same is true for the insource-outsource relationships. However, the compositions of clusters regarding the onshore-offshore relationships were found significantly different, showing that the offshore participants have stronger tendency to cluster together. The correlation analysis between the actual and ideal situation gave interesting results, for example the consistency of insource stakeholders regarding the majority of the goals and the inconsistency of architects. Finally, the graphical analysis provided useful tools for visualizing agreements between stakeholders. The main result is the fact that it is possible to formulate a statistical framework that could help in analyzing these types of situation. Other results would of course be expected with a different data set, although applying the same framework.

The paper presents in detail the methodologies followed and discusses the results of their application to the specific data. It is important to note that the whole approach is generic in the sense that it can be applied to any other similar situation involving prioritization data. The paper essentially presents an experienced-based framework, having as starting point a multivariate analysis which considers the views of all stakeholders on all goals and forms clusters which are then analyzed with respect to features either "internal" (used for clustering) or "external" (others than the ones used for clustering).

The study is exploratory and shows the need for further research to discover the causes of diversity in prioritization of goals among stakeholders. These causes are not necessarily registered as data and can have their origins in various cultural or ethical issues or even in personal beliefs and attitudes. At this

point it is essential to emphasize that the proposed framework does not aim to substitute any other traditional or even novel methods of team management. It is intended to be a supplementary exploratory tool that quantifies, systematizes and investigates human preference towards certain goals. It can be used by decision makers either to confirm what is already known from personal experience or to investigate further and in detail unanticipated results.

The paper is structured as follows: Section 2 provides an outline of the related work. Section 3 presents the framework as a set of statistical tools addressing specific research questions, along with their basic principles. Section 4 describes the dataset used for the illustration of the framework. Section 5 presents the results from the application of statistical analysis. We discuss the findings of the paper and the threats to the validity of the study in Section 6. Finally, the conclusions from the analysis are presented in Section 7 where we provide some directions for future work.

2. RELATED WORK

The present paper is an extended work of the paper presented by Chatzipetrou et al. (2011) at the 37th Euromicro conference on Software Engineering and Advanced Applications (SEAA) and is a continuation of the works published in Barney and Wohlin, 2009b; Barney and Wohlin, 2010; Barney et al. 2011. The paper by Chatzipetrou et al. (2011) presented only a portion of the methods and results, and only for the actual situation. In the present work we present a holistic framework of statistical methods for the analysis of goal alignment among stakeholders along with the results of their application to both actual and ideal situations.

At this point it is necessary to clarify what is meant by “alignment”. Although the term can be perceived in several different ways, for example as communication, collaboration, efficient management or leadership of teams in a disciplined manner, we do not refer to these concepts. For this study, alignment means that: (a) there are different individuals having similar prioritizations of a predefined number of goals; (b) the prioritization of each individual can be expressed in the form of a ranking or voting scheme which is able to capture and quantify the individual’s preferences; (c) each individual prioritizes the predefined goals independently of the other individuals and (d) alignment can be expressed as a numerical function calculated by the quantified prioritizations of two or more individuals. According to the aforementioned definition, alignment is expressed as a similarity function between individuals whose prioritizations are represented by numerical vectors.

Barney and Wohlin (2009b) have proposed a research methodology to analyze prioritization data of software quality goals. The method was applied in a case study concerning specific products developed by a world leading company in telecommunication. Based on Theory-W (Boehm and Ross, 1989), they identified the success critical stakeholder groups, and by using an ISO 9126 (2001)-based quality model,

they developed a questionnaire (Barney and Wohlin, 2008), which was filled by the representatives of each of the identified success-critical stakeholder groups. At this point they used the method of Hierarchical Cumulative Voting (HCV) (Berander and Jönsson, 2006a, 2006b) to elicit the priorities given to various goals. Finally, they performed exploratory analysis on the data trying to determine the degree to which the key stakeholders are aligned regarding how they perceive quality in two different situations (actual and ideal).

The importance of stakeholder alignment has been recognized as crucial for the success of a project (Chan, 2002; Chan and Reich, 2007). Barney and Wohlin (2010) introduced an empirical study of alignment between key stakeholders presenting a case study that aimed to compare two groups of stakeholders: internal to the development organization, and from outsourcing relationships. In Barney et al. (2011), in a single case study, and in a following work (Barney et al., 2014), in three different case studies which took place in three different continents, the authors studied the level of alignment between groups using a specific numerical metric.

It should be noted that the general topic of alignment, in the sense of collaboration, has been extensively studied in a number of areas. For example, in the product development process, the managers affect the speed and productivity and the senior managers affect the product effectiveness, however the overall effective process is based on the successful collaboration between the different groups of stakeholders (Brown and Eisenhardt, 1995; McDonough and Spital, 2003). In particular in the global software development literature and foremost in the area of requirements engineering, studies had showed that stakeholders need effective knowledge acquisition and sharing as well as relationship building practices in order to overcome the significant cultural, time zone, and organizational challenges (Damian, 2007; Damian and Zowghi, 2003). All of the above depend on the members' role (Espinosa et al., 2007).

The framework we propose in the present paper is based on data obtained by the HCV method. The original method of Cumulative Voting (CV) (or the hundred dollar test) is described in Leffingwell and Widrig (2003) as a simple, straightforward and intuitively appealing voting scheme where each stakeholder is given a fixed amount (e.g. 100, 1000 or 10000) of imaginary units (for example monetary) that can use for voting in favor of the most important items (issues, goals, requirements etc.). In this way, the amount of money assigned to an item represents the respondent's relative preference (and therefore prioritization) in relation to the other items. The points can be distributed according to the stakeholder's opinion. Each stakeholder is free to put the whole amount on only one item of dominating importance or to distribute equally the amount to several or even to all of the issues. The multivariate data obtained from CV are of special type, since the values of all variables have a constant sum. Statistical analysis of CV data was also performed in Rovegård et al. (2008) and in Chatzipetrou et al. (2010).

As already mentioned, GSD has become nowadays a common practice in software development with a number of benefits and risks (Conchúir et al. 2009). Many studies cite the ability to reduce development costs using GSD techniques and to focus on strategic business functions (Conchúir et al. 2009; Morgan, 2004; Sahay et al., 2003). Thus, it is of particular interest to study a statistical framework's usefulness in this context where many factors could affect the likelihood of being aligned when it comes to for example quality goals.

While the value of stakeholder alignment is clear, there is a complex myriad of factors that can stand in the way of software development organizations achieving this goal. Understanding the level of alignment and reasons for the level of alignment with appropriately collected data and statistical analysis is of value to organizations, as this information can be used to support improvement initiatives. A recent study (Pinjani and Palvia, 2013) is an excellent example of how statistics can be used for understanding the different facets of diversity present in global virtual teams.

3. THE STATISTICAL FRAMEWORK

The present paper explores in detail the alignment through a statistical framework which combines cluster analysis, cluster evaluation, correlation analysis and graphical representation of the closest alignment. The framework follows and addresses a structured sequence of research questions, which start from investigation of general aspects (searching of clusters) and continue to more refined issues (study of clusters, consistency and study of alignment at an individual level).

The basic research question is:

RQ0: Is it possible to use a statistical framework to give valuable insights regarding the alignment of key stakeholders towards specific goals?

The following five research questions are generic and they were studied in the specific case to illustrate the statistical framework and its usefulness:

- RQ01: How are the stakeholders clustered together according to their alignment in prioritization of goals, especially when this prioritization is expressed for different situations?
- RQ02: Are there differences in the composition of clusters of stakeholders regarding their known characteristics (e.g. role, location, relationship with the company etc.)?
- RQ03: In which of the goals is there a significantly different prioritization between the clusters of aligned stakeholders?
- RQ04: Is there a consistency in the views of stakeholders when they prioritize for different situations? Which special groups of stakeholders are consistent?
- RQ05: Given that each stakeholder has closest alignment with some other stakeholder; can we represent this relationship for the whole dataset in a simple graphical way?

All of the abovementioned research questions are essentially an attempt to explore and understand the structure of agreement in the stakeholders' views under different perspectives or situations. In the following subsections we discuss the statistical methods which address the aforementioned research questions, aiming to comprise a generic methodology that first discovers clusters of aligned stakeholders and then explores the composition of the discovered clusters with respect to the various inherent groups. The methodology also involves comparative analysis of the goals within each cluster and analysis of consistency of views under different situations. Furthermore, a graphical representation is suggested for the closest alignment relationship. For details of the statistical tests proposed, we refer to Sheskin (2004).

3.1 Hierarchical Cluster Analysis

Hierarchical cluster analysis (HCA) (Johnson and Wichern, 2002) is a multivariate statistical procedure that attempts to identify relatively homogeneous groups of individual cases. The grouping (or clustering) is based on values of common variables which are used to describe all individuals. In our case, HCA is used to find groups of stakeholders based on their prioritizations of goals. This analysis uses an agglomerative algorithm. The algorithm performs a stepwise procedure that starts with each stakeholder in a separate cluster. In each step of the algorithm the previously formed clusters are combined into larger clusters, until all stakeholders are members of only one cluster. In the course of the algorithm, the clusters are combined on the basis of the stakeholders' agreement in their prioritizations of goals.

Agreements or disagreements are elements of a proximity matrix, which contains the values of a measure expressing stakeholders' similarity or dissimilarity. The principal advantage of the hierarchical procedure is that a tree-like hierarchy is structured so as to see in detail the consecutive connections of the stakeholders as the algorithm proceeds.

The agglomerative algorithm in HCA is based on distances between the stakeholders as individuals and between the clusters formed. In our application we used the implementation of HCA in SPSS (Field, 2009), which requires an initial distance matrix between individuals and a criterion to combine clusters. In order to obtain a meaningful result, we tried several combinations and we finally used as initial distance matrix the one that uses as distances between stakeholders the chi-square metric (Rovegård et al., 2008), suitable for counts, as in our case the amounts of monetary units assigned to each goal, and the Ward method (Johnson and Wichern, 2002) as a criterion for combining clusters.

Although in literature there is a plethora of clustering algorithms, especially for large datasets, in our framework we suggest to address RQ01 with HCA since the method provides a step-by-step graphical representation of the whole procedure which is very informative about the way stakeholders are aligned.

Indeed, the most important output of HCA is the dendrogram, i.e. a tree diagram, which is a visual representation of the steps of the agglomerative algorithm. It shows how the stakeholders are combined at each step and also visualizes the rescaled distances with connected lines. Dendrograms can be used to assess the cohesiveness of the clusters formed and can signify the appropriate number of clusters. In our analysis we use dendrograms to show how stakeholders belonging to various roles and inherent groups are clustered together.

3.2 Methods to analyze clusters

RQ02 and RQ03 require analysis of the stakeholders' characteristics within and between clusters. In order to test which goals were prioritized significantly different between clusters, we suggest the use of non-parametric tests for independent samples. In the case of two clusters, the most appropriate is the Mann-Whitney test, while for more than two clusters the Kruskal-Wallis test can be used. Also, as a graphical tool for accompanying the analysis, box-plots are especially informative for depicting differences.

The compositions of the clusters with respect to the inherent groupings of stakeholders (e.g. different roles, relation with the company, etc.) can be studied by contingency tables and the accompanying chi-square test.

3.3 Consistency of prioritizations for different situations

As we already mentioned, for the needs of the research, the stakeholders may prioritize the goals for different situations or perspectives. It is therefore necessary to test the consistency of these prioritizations (RQ04). Consistency can be measured by the non-parametric Spearman rank correlation coefficient ranging from -1 to 1. A value of the Spearman coefficient close to one, shows high level of consistency in the views of all or of a group of stakeholders when they prioritize under two situations. On the other hand, a value close to zero shows inconsistency of stakeholders' prioritizations for the situations under study. A value close to -1 would be especially interesting as it would show an overturned prioritization, i.e. systematic inconsistency. The corresponding p-values of the Spearman coefficients indicate significant relationships (positive or negative) between the two situations. So, RQ04 is essentially addressed by correlation analysis. In our illustrative application, which is described in Section 4, we test for consistency the prioritization under the actual and the ideal situation.

3.4 The closest alignment approach

In order to investigate further the issue of alignment, beyond clustering and at an individual level (RQ05), we considered the possibility to represent graphically the whole dataset in a manner that brings out the way each one of the stakeholders is "connected" to the person who has the closest alignment in views of prioritization. For this purpose, we suggest the use of a graph, i.e. a set of nodes and directed edges (arrows). Each node stands for a stakeholder while an arrow starting from node A and ending to node B shows that stakeholder B is the person with the closest prioritization view to stakeholder A. The closest alignment is decided by finding the minimum chi-square distance between a specific stakeholder and all the others. Note that the relation is not symmetric. Therefore, stakeholder B can be the closest alignment of stakeholder A, but A may not be the closest alignment of B. The overall graph shows a very interesting structure of the dataset which can be used to identify "isolated" or "concentrative" stakeholders' views.

4. ILLUSTRATION OF THE FRAMEWORK BY A CASE STUDY

The primary purpose of the case study presented in this section is to illustrate the possibilities and usefulness of the statistical framework. It should be emphasized that the data used here have been obtained from responses of employees in a real working environment. The participants do not comprise a random sample from a large population. The categories represented neither have resulted from a balanced experimental study nor survey design. This fact may be seen as an inherent weakness of the data, in the sense that they cannot be used for inferences and generalizations to large populations or to answer questions like RQ02 in general, but only in the specific case. However, they are utilized here for illustrational purposes in order to show how the prioritization views of different people working for the same goals can be quantified and systematically analyzed under a unified framework which is able to provide a holistic map of their alignment with respect to specific goals.

The case study as such concerns one major product developed by a world leading company in telecommunications, providing a wide range of products and services. These are developed and sold as generic solutions, although customized versions of the product are also developed for key customers. This product is a leading solution for the market and from now on is referred simply as "Product" throughout this paper for reasons of confidentiality.

Stakeholder groups are classified in two dimensions (Šmite et al., 2014):

1) *Insourcing/Outsourcing*

- *Insourcing (IN)*, defines that the work is undertaken by employees internal to the organization.
- *Outsourcing (OUT)*, defines stakeholders employed by a consultancy company or any other subcontractor to work for the organization.

2) Onshore/Offshore

- *Onshore (ON)*, defines that the work is undertaken in the home country of the organization (in our case Sweden).
- *Offshore (OFF)*, defines that the work is conducted outside the boundaries of Sweden and specifically in India (in our case study).

Furthermore, seven roles were identified according to the responsibilities each stakeholder has in the organization. These roles are:

- *Architects (ARCH)* are responsible for the overall product design, assigning requirements to modules of the product.
- *Developers (DEV)* are responsible for the implementation of requirements.
- *Product Support (PS)* provides support to users and product owners at customer sites.
- *Project Managers (PM)* are responsible for planning and executing projects aligned with the priorities of the strategic product management.
- *Strategic Product Managers (SPM)* have the strategic product responsibility and decide the overall product development direction.
- *Tactical Product Managers (TPM)* support the strategic product management with expert knowledge of the systems and their architecture. It is also responsible for providing analysis of pre-project requirements in the form of feasibility, impact and technical dependencies.
- *Testers (TSR)* are responsible for the verification and validation of requirements.

The distribution of the stakeholders within the inherent identified groupings is shown in Table I.

Each stakeholder prioritized 24 goals, all related to Software Product Management. The goals were prioritized using Hierarchical Cumulative Voting (HCV) in two levels.

The process of defining those goals was a collaborative exercise described in detail in (Barney and Wohlin 2009b). The procedure involved academic and industrial perspectives. The starting point was the ISO 9126 standard (2001), which then was complemented with a company specific view of important software quality goals. A list of goals was defined specifically for the products of the organization, maximizing the relevance to this industrial partner and the possibilities for using the results to support improvements within the company. However, just as some goals can be more important than others, there are other aspects of the development process that compete with the implementation of software quality.

The authors (Barney and Wohlin, 2009a) wrote preliminary definitions for these terms and a workshop was held within the organization to review and refine the terms defining software quality. The aim was to ensure the final list of terms and definitions would be complete, meaningful and useful to the organization. Finally the model was split into three categories – the ISO 9126 (2001) goals relating to *functionality*, *system properties* and *project management* and each of them contains a number of lower level goals.

Specifically the three goals of the higher level are:

- A is related to *Features* (contains 4 goals on the lower level),
- B is related to *System Properties* (contains 18 goals on the lower level),
- C is related to *Project Management* (contains 2 goals on the lower level),

A description of the goals in both levels (Higher and Lower) is shown in Table II.

Role	Abbrev	TOTAL	IN (Insource)	OUT (Outsource)	ON (Onshore)	OFF (Offshore)
Architects	ARCH	5	5	0	0	5
Developers	DEV	22	15	7	15	7
Product Support	PS	2	2	0	0	2
Project Managers	PM	7	7	0	4	3
Strategic Product Managers	SPM	9	9	0	7	2
Tactical Product Managers	TPM	10	10	0	9	1
Testers	TSR	10	10	0	3	7
TOTAL		65	58	7	38	27

TABLE I. CATEGORIZATION OF STAKEHOLDERS

All stakeholders were asked to prioritize the goals using Hierarchical Cumulative Voting (HCV) from two different perspectives. Specifically, they prioritized according to their views of the actual situation and how they believe the priorities *should be*, i.e. how they think the organization should be operating according to their perceived *ideal* situation.

For each situation a stakeholder distributed 1000 imaginary monetary units, first to the three goals of the higher level: Features (A), System Properties (B) and Project Management (C). Next, for each one of these higher-level goals, each stakeholder distributed again 1000 monetary units to its goals at the lower level. The more money an individual assigns to a goal, the higher the prioritization is. The data obtained by HCV were transformed to simple CV results by a procedure described by Berander and Jönsson (2006b), which takes into account the number of lower level goals belonging to each one of the higher levels.

Higher Level	A: Features		B: System Properties		C: Project Management	
Lower Level	A1	Suitability	B1	Security	C1	Time
	A2	Accuracy	B2	Maturity	C2	Cost
	A3	Interoperability	B3	Recoverability		
	A4	Certification/ Standardization/ Compliance	B4	Containment/ISP/Fault tolerance		
			B5	Learnability		
			B6	Understandability		
			B7	Operability		
			B8	Time behavior		
			B9	Resource behavior		
			B10	Robustness/Stability		
			B11	Performance Management/Statistics		
				B12	Analyzability	
			B13	Changeability		
			B14	Testability		
			B15	Installability		
			B16	Upgradability/Replaceability		
			B17	Configurability/Product Customizability/Adaptability		
				B18	Scalability	

TABLE II. GOALS IN TWO LEVELS

5. RESULTS FROM USING THE STATISTICAL FRAMEWORK

In this section we present the results of the analysis of the prioritized data for both the actual situation and for the perceived ideal situation and separately for: (a) the different roles of stakeholders, (b) the insource/outsource grouping and (c) the onshore/offshore arrangement.

5.1 Cluster analysis results

The hierarchical cluster analysis was applied to a prioritization data matrix consisting of 65 rows and 24 columns. Each row represents an individual stakeholder while the 24 columns contain the transformed prioritizations of the goals described in Section 4. Since the original data were obtained by a HCV in two levels, the 24 variables used in the analysis were transformations by the Berander and Jönsson (2006b) method.

The results from the hierarchical cluster analysis show that although there are some local groupings at the lower levels, we can distinguish two major clusters for both the actual situation and the ideal

situation. For the actual situation, Cluster 1 contains 48 stakeholders and Cluster 2 contains 17 stakeholders. For the ideal situation, Cluster 1 contains 44 stakeholders while Cluster 2 contains 21 stakeholders.

		Cluster results for ideal		Total
		Cluster 1	Cluster 2	
Cluster results for actual	Cluster 1	37	11	48
	Cluster 2	7	10	17
Total		44	21	65

TABLE III. CLUSTER RESULTS FOR ACTUAL AND IDEAL SITUATION

Table III is a contingency table, showing the common distribution of stakeholders according to the cluster they belong under the perceived situation that they prioritize. As we can see, 37 stakeholders are common in the large clusters of both situations while 10 of them are common in the small clusters. The chi-square test for independence between the two cluster formations has significance $p=0.007$, indicating that formations of the two clusters under the two situations are not independent. The result implies that stakeholders who are aligned in their prioritization of the actual situation are most probably aligned in their prioritization of the ideal situation, i.e. when a group of people agree in their views of goals of the actual situation their agreement is most probably maintained when they describe the ideal situation. Of course, if we analyze further the common distribution of Table III, besides the majority of stakeholders who remain stable in the same cluster in both situations, there are stakeholders who change cluster when they prioritize under different situation. It is quite interesting to inspect the roles and the characteristics of the stakeholders who change cluster.

Testers (TSR) seem to be the most stable stakeholders. All TSRs stakeholders, apart from one, remain in the same cluster in both situations. Developers (DEV) also seem to be reasonably stable. Only 20% (5 out of 22) change cluster when they prioritize under a different situation. On the other hand, Architects (ARCH) and people in Product Support (PS) seem to be the role categories changing cluster the most. Only 3 out of 5 Architects (60%) remain stable in both situations. The rest 40% (2 out of 5) change cluster when they prioritize under different situation. Regarding the Product Support (PS) category, there are only two stakeholders holding this role. Those two stakeholders are not aligned in their prioritization of the two situations. Only one of them remains in the same cluster.

The entire clustering schemes for the actual and the ideal situation are shown in the dendrograms of Fig. 1a) and b) respectively. The stakeholders are labeled with their role and inherent groupings in order

to show how they are combined in clusters. Table IV shows the distributions of roles within derived clusters.

In the actual situation, it is noticeable that all Testers (TSR) and both of the Product Support (PS) stakeholders are members of the large cluster. In general, there are differences in the compositions of the clusters. For example, Developers (DEV) have the largest percentage in Cluster 1, while in Cluster 2 Tactical Product Managers (TPM) hold the largest percentage. Moreover, we can see that although some stakeholders with the same role (especially Developers and Testers) are combined at early stages of the algorithm, showing similarity in prioritizations, in general there is no significant statistical evidence that the clustering is associated with the roles (the chi-square test gives $p > 0.05$). This means that if we try to explain the alignment of various stakeholders' in terms of their prioritization, we have to discover other factors or characteristics, apart from their role, which may affect the agreement. The role itself is a very weak criterion for judging the stakeholders' preferences. This is also true for the ideal situation where even the Testers and the Product Support roles are split.

The compositions of clusters regarding the insource/outsource grouping shows that the seven outsource stakeholders (which are all in an onshore relationship) are clustered together in the first cluster but in general the connections among them are not strong. As we can see, they are grouped together at later steps of the algorithm. Also, the appearance of all of them in the first larger cluster is not statistically significant, i.e. there is no significant statistical evidence ($p > 0.05$) that their clustering depends on the insource/outsource characteristic. For the ideal situation, the compositions of the two clusters are quite similar, so it is clear that the clustering is not associated to insourcing/outourcing characteristic. Therefore, the insource/outsource arrangement of stakeholders does not offer any significant information to explain their alignment in prioritization, which may indicate either that the company has managed to communicate their priorities in relation to quality or that the company being outsourced to has good domain knowledge.

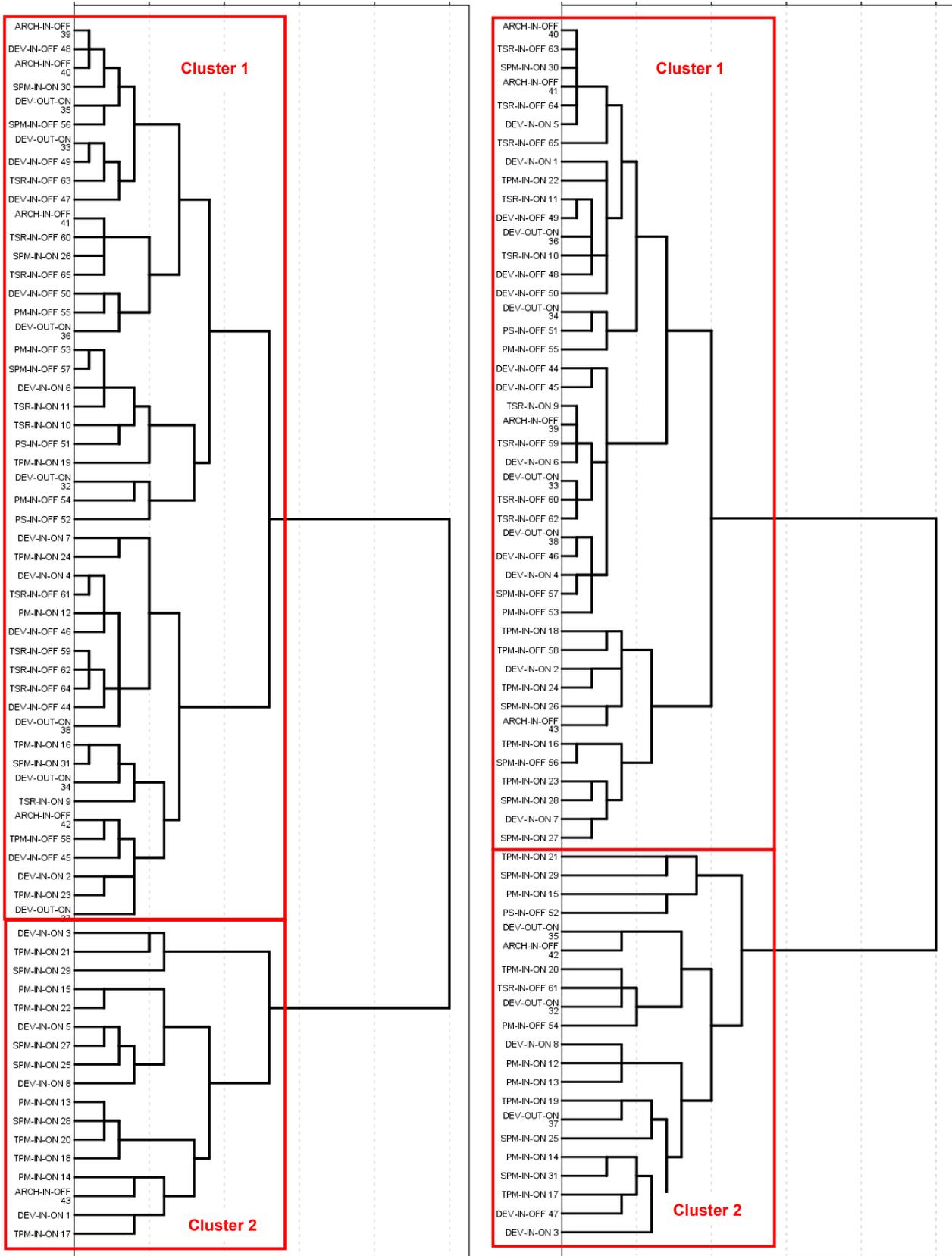


Figure 1. a) Clustering of stakeholders for the actual situation

b) Clustering of stakeholders for the situation ideal

<i>Actual Situation</i>	<i>Role</i>	<i>ARCH</i>	<i>DEV</i>	<i>PM</i>	<i>PS</i>	<i>SPM</i>	<i>TPM</i>	<i>TSR</i>	<i>Total</i>
Cluster 1	Count (%)	4(8.3%)	18(37.5%)	4(8.3%)	2(4.2%)	5(10.4%)	5(10.4%)	10(20.8%)	48 (100%)
Cluster 2	Count (%)	1(5.9%)	4(23.5%)	3(17.6%)		4(23.5%)	5(29.4%)		17(100%)
Total	Count (%)	5(7.7%)	22(33.8%)	7(10.8%)	2(3.1%)	9(13.8%)	10(15.4%)	10(15.4%)	65(100%)
	<i>Insource/ Outsource</i>	<i>Insource</i>			<i>Outsource</i>				<i>Total</i>
Cluster 1	Count (%)	41(85.4%)			7(14.6%)				48(100%)
Cluster 2	Count (%)	17(100%)							17(100%)
Total	Count (%)	58(89.2%)			7(10.8%)				65(100%)
	<i>Onshore/ Offshore</i>	<i>Offshore</i>			<i>Onshore</i>				<i>Total</i>
Cluster 1	Count (%)	26(54.2%)			22(45.8%)				48(100%)
Cluster 2	Count (%)	1(5.9%)			16(94.1%)				17(100%)
Total	Count (%)	27(41.5%)			38(58.5%)				65(100.0%)
<i>Ideal Situation</i>	<i>Role</i>	<i>ARCH</i>	<i>DEV</i>	<i>PM</i>	<i>PS</i>	<i>SPM</i>	<i>TPM</i>	<i>TSR</i>	<i>Total</i>
Cluster 1	Count (%)	4(9.1%)	16(36.4%)	2(4.5%)	1(2.3%)	6(13.6%)	6(13.6%)	9(20.5%)	44 (100%)
Cluster 2	Count (%)	1(4.8%)	6(28.6%)	5(23.8%)	1(4.8%)	3(14.3%)	4(19.0%)	1(4.8%)	21(100%)
Total	Count (%)	5(7.7%)	22(33.8%)	7(10.8%)	2(3.1%)	9(13.8%)	10(15.4%)	10(15.4%)	65(100%)
	<i>Insource/ Outsource</i>	<i>Insource</i>			<i>Outsource</i>				<i>Total</i>
Cluster 1	Count (%)	40(90.9.4%)			4(9.1%)				44(100%)
Cluster 2	Count (%)	18(85.7%)			3(14.3%)				21(100%)
Total	Count (%)	58(89.2%)			7(10.8%)				65(100%)
	<i>Onshore/ Offshore</i>	<i>Offshore</i>			<i>Onshore</i>				<i>Total</i>
Cluster 1	Count (%)	22(50.0%)			22(50.0%)				44(100%)
Cluster 2	Count (%)	5(23.8%)			16(76.2%)				21(100%)
Total	Count (%)	27(41.5%)			38(58.5%)				65(100.0%)

TABLE IV. COMPOSITION OF CLUSTERS WITH RESPECT TO THE INITIAL GROUPS

Finally, regarding the onshore/offshore grouping, it is clear that the compositions of clusters are different. It is interesting here to see that the second cluster contains mainly onshore stakeholders. Almost all of the offshore stakeholders are grouped in the first cluster at the early stages of the algorithm. Here, the chi-square statistic shows significant association between the clustering and the initial onshore/offshore grouping ($p=0.001$). The same is true for the ideal situation. Here the chi-square test gives $p=0.045$ showing that the dependence is significant though it seems to be not as strong as for the actual situation.

The results in both perspectives, actual and ideal, show that the alignment of stakeholders is affected by the onshore-offshore relationship and this seems to be an important factor for explaining agreements, which deserves further investigation. Thus, the cluster analysis proposed as part of the statistical framework gives an opportunity to identify differences and similarities in views among the stakeholders.

5.2 Evaluation of the clusters

In an effort to identify and compare the profiles of the members forming the derived clusters, we tried to depict how the stakeholders in each one of the clusters prioritize the goals. Having in mind that clusters are essentially groups of people with some degree of coherence in their views, we are interested in comparing the distribution of prioritizations of goals in each one of them. In Fig. 2 - 4 we can see respectively how the stakeholders in each cluster prioritize the goals related to Features (A1-A4), System Properties (B1-B18) and Project Management (C1, C2) for the actual (a) and ideal situation (b) respectively. The boxplots show the distribution of the prioritization values each goal received, separately in each cluster. Both clusters are represented in the same figure for easy comparison.

It is noticeable that the prioritizations of A1 (suitability) in Cluster 2 have a large variation. In general, all goals of Features are prioritized higher in Cluster 2. On the contrary, the goals of System Priorities receive lower ranking in Cluster 2. Regarding the two goals of Project Management, Time seems to be more important for Cluster 2 while Cost seems to be equally important for both clusters. Another interesting point is that all goals are prioritized more uniformly in Cluster 1 than in Cluster 2 where we generally have larger discrepancies.

Regarding the results for the ideal situation, some of them are quite similar to the actual situation, but there are also some differences. Members of Cluster 1 prioritize the goals in a more uniform manner; A1 (suitability) seems to be more important for members of Cluster 2 while the other goals of Features seem to be equally important for both clusters. Some of the goals of System Properties are most important for Cluster 1 while others are equally important. Both goals of Project Management seem to be equally important for both clusters. In general, in Cluster 2 we can observe larger deviations.

Furthermore, we applied the non-parametric Mann-Whitney test on each one of the goals separately, to see which ones have significance differences between the two clusters. In this way we can characterize the clusters according to their tendency to prioritize more or less certain goals. The results of the test are shown in Table V where we report the p-value of the test for each goal and for both the actual and ideal perspective. The goals with significant differences between clusters (p-value) are written in bold. These are: for the actual situation, goals A1, A2, A3, B3, B4, B5, B6, B7, B12, B13, B14, B15, B16, B18 and C1 and for the ideal situation, goals A1, B2, B3, B6, B7, B8, B10, B13 and B14 have statistically significant difference between clusters.

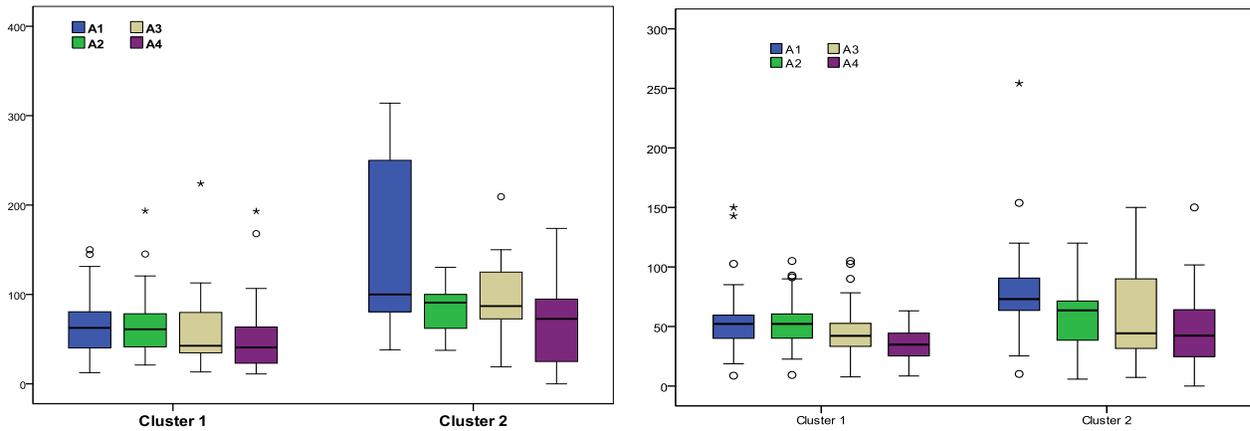


Figure 2. Goals related to Features (A1-A4) for a) actual situation b) ideal situation

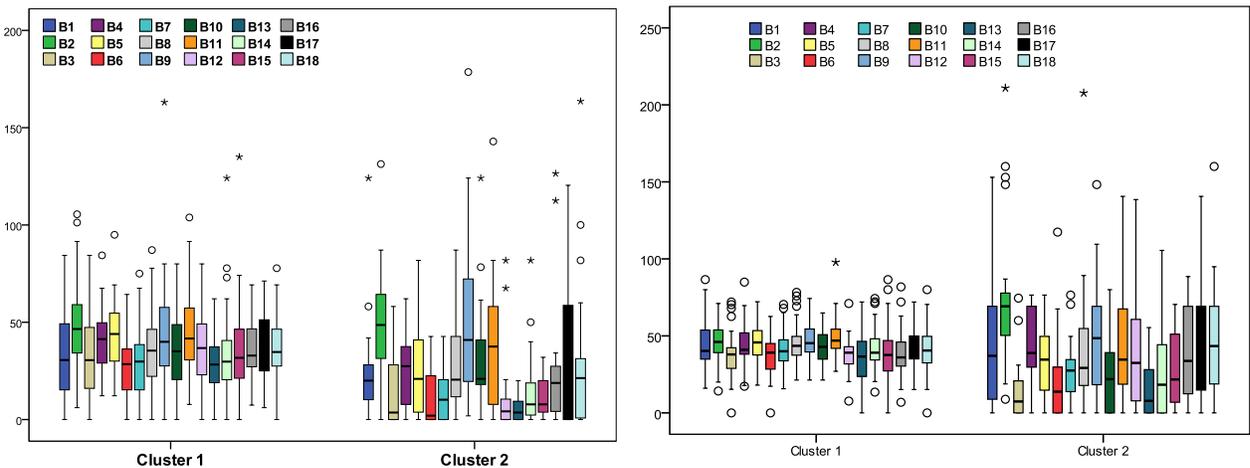


Figure 3. Goals related to System Properties (B1-B18) for a) actual situation b) ideal situation

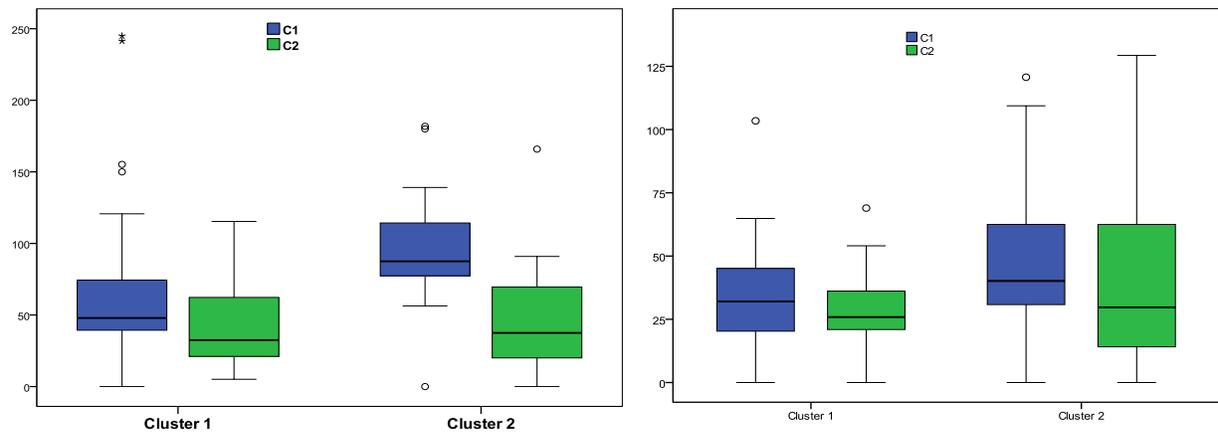


Figure 4. Goals related to Project Management (C1-C2) for a) actual situation b) ideal situation

Software goals	Actual	Ideal	Software goals	Actual	Ideal
Suitability (A1)	<0.001	0.004	Resource behavior (B9)	0.976	0.523
Accuracy (A2)	0.008	0.428	Robustness/Stability (B10)	0.289	0.005
Interoperability (A3)	0.003	0.381	Performance Management/Statistics (B11)	0.263	0.483
Compliance (A4)	0.177	0.214	Analyzability (B12)	<0.001	0.340
Security (B1)	0.092	0.833	Changeability (B13)	<0.000	<0.001
Maturity (B2)	0.676	0.001	Testability (B14)	0.002	0.007
Recoverability (B3)	0.003	<0.001	Installability (B15)	<0.001	0.060
Fault tolerance (B4)	0.008	0.710	Upgradability/Replaceability (B16)	0.001	0.408
Learnability (B5)	0.002	0.052	Configurability/Product Customizability/Adaptability (B17)	0.156	0.369
Understandability (B6)	0.002	<0.001	Scalability (B18)	0.030	0.752
Operability (B7)	0.002	0.001	Time (C1)	0.001	0.630
Time Behavior (B8)	0.210	0.041	Cost (C2)	0.893	0.424

TABLE V. RESULTS OF THE MANN-WHITNEY TEST FOR THE DIFFERENCES BETWEEN CLUSTERS

5.3 Correlation results

The correlation analysis described in this section tries to identify the consistency of prioritization under the two different perspectives: the actual situation against the ideal situation. Having in mind that our data set consists of different inherit stakeholder groups, we are interested in testing the significance of this consistency separately for each one of these groups.

For this purpose, we calculated the Spearman rank correlation coefficients between the two different perspectives, for each goal and separately for each stakeholder group in order to investigate the level of stakeholders' agreement between the two situations, actual and ideal. So the Spearman rank correlation coefficients are calculated separately for the entire dataset of stakeholders (overall) and separately for (a) the insource/outsource grouping, (b) the onshore/offshore grouping and (c) the role. The results are shown in Table VI where for each group the Spearman correlation coefficient is accompanied by an indication of significance. Specifically, bold coefficients with “*”, show statistically significant correlation at 0.05 significance level, while bold coefficients with “***”, show stronger statistically significant correlation at 0.01 significant level.

ACTUAL VS IDEAL	OVER ALL	INSOURCE/		ONSHORE/		ROLE					
		OUTSOURCE		OFFSHORE		DEV	TSR	PM	TPM	SPM	ARCH
		IN	OUT	ON	OFF						
A1	0.224	0.268*	-0.143	0.207	0.126	-0.008	0.382	0.179	-0.285	0.867**	-0.500
A2	0.188	0.117	0.857*	0.233	0.041	0.327	0.733*	-0.214	0.030	0.243	-0.100
A3	0.306*	0.384**	-0.929**	0.341*	0.269	0.419	-0.006	0.714	-0.030	0.133	0.000
A4	0.411**	0.441**	0.464	0.547**	0.041	0.292	0.212	0.714	0.780**	0.783*	-0.400
B1	0.319**	0.252	0.857*	0.293	0.333	0.460*	0.309	0.745	-0.104	0.351	-0.051
B2	0.326**	0.249	0.679	0.370*	0.324	0.225	0.479	0.679	0.389	0.243	0.800
B3	0.455**	0.481**	0.000	0.521**	0.115	0.168	0.273	1.000**	0.392	0.717*	-0.900*
B4	0.116	0.111	-0.072	0.034	0.204	-0.302	0.055	0.107	0.006	0.828**	0.400
B5	0.218	0.241	0.000	0.181	0.255	0.107	-0.073	-0.036	0.195	0.444	-0.300
B6	0.476**	0.533**	-0.234	0.499**	0.446*	0.009	0.6	0.711	0.018	0.933**	-0.300
B7	0.349**	0.377**	-0.036	0.248	0.222	-0.004	0.503	0.482	-0.052	0.552	-0.718
B8	0.324**	0.321*	0.45	0.32*	0.337	0.233	0.236	0.786*	-0.055	0.9**	-0.400
B9	0.385**	0.312*	0.929**	0.436**	0.359	0.422*	0.406	0.709	-0.055	0.301	-0.500
B10	0.436**	0.402**	0.643	0.369*	0.505**	0.46*	0.721*	0.679	-0.333	0.828**	-0.462
B11	0.050	0.065	0.143	-0.178	0.453*	0.13	0.042	-0.214	0.018	0.217	0.900*
B12	0.38**	0.397**	0.198	0.155	0.662**	0.276	0.321	0.964**	0.024	-0.218	0.800
B13	0.300**	0.341**	-0.321	0.162	0.485**	-0.285	0.503	0.676	0.411	0.243	0.800
B14	0.306*	0.311*	0.321	0.232	0.481**	0.266	0.626	0.972**	-0.122	0.633	-0.100
B15	0.394*	0.400**	0.018	0.304	0.435*	0.101	0.576	0.929**	-0.137	0.667*	0.400
B16	0.370**	0.365**	0.607	0.391*	0.151	0.338	0.28	0.865**	-0.219	0.817**	0.300
B17	0.297**	0.302*	0.107	0.383*	0.058	0.303	-0.055	0.571	0.122	0.845**	-0.500
B18	0.265*	0.267*	0.321	0.461**	-0.104	0.453*	-0.127	-0.775*	0.523	0.517	-0.600
C1	0.131	0.148	0.491	0.237	-0.028	0.018	-0.200	0.286	0.644*	0.167	-0.205
C2	0.308*	0.245	0.468	0.407*	0.162	0.533*	0.333	0.500	0.267	-0.233	-0.821

TABLE VI. SPEARMAN RANK CORRELATION COEFFICIENS

Note that there are no coefficients for the PS stakeholders, since there are just two of them. Generally and without any inherit stakeholders' grouping, at the 0.05 level there is significant correlation and therefore consistency in the views with respect to two different perceptions for a number of goals. Also, for various groups there are interesting consistencies.

More specifically, regarding the whole dataset (OVERALL) and the groups:

- All stakeholders (OVERALL): For the whole dataset, there seems to be statistically significant, but not so strong consistency, in the preferences of the majority of goals. The only goals where the consistency *is not significant* are: A1, A2, B4, B5, B11 and C1. The Spearman coefficient even for significant correlations is always less than 0.5.

- Insource (IN): The stakeholders of this group show significant consistency for A1, A3, A4, B3, B6, B7, B8, B9, B10, B12, B13, B14, B15, B16, B17 and B18. The strongest correlation is for B6.
- Outsource (OUT): The outsource stakeholders have consistency in their views under the actual situation and the ideal situation only for A2, A3, B1 and B9.
- Onshore (ON): The onshore stakeholders have statistically significant consistency in A3, A4, B2, B3, B6, B8, B9, B10, B16, B17, B18 and C2.
- Offshore (OFF): The offshore group has significant consistency in B6, B10, B11, B12, B13, B14 and B15.
- Developers (DEV): The stakeholders in this role have consistency in B1, B9, B10, B18 and C2.
- Testers (TSR): There is significant consistency only in A2 and B10.
- Project Managers (PM): There is significant consistency in B3, B8, B12, B14, B15 and B16. Notable is the statistically significant negative correlation in B18 showing significant inconsistency.
- Tactical Product Managers (TPM): There is significant consistency only in A4 and C1. Actually, this is the only group where C1 has consistency.
- Strategic Product Managers (SPM): There is significant consistency in A1, A4, B3, B4, B6, B8, B10, B15, B16 and B17.
- Architects (ARCH): The group shows significant consistency only in B11, while notable is the negative correlation in B11 which shows significant inconsistency.

The only goal with no significant consistency, either for all stakeholders or for any of the groups, is B5. In general, the proposed framework allows the identification of this type of consistencies for any grouping. The correlation analysis provides more in-depth understanding regarding the prioritizations with respect to actual and ideal situations.

5.4 Results from the closest alignment approach

In this Section we present two directed graphs, one for the actual situation (Fig. 5) and one for the ideal situation (Fig. 6). Each node stands for a stakeholder described by the groups each one belongs to and a serial number. The arrow from A to B shows that B is the nearest neighbor of A, i.e. the closest prioritization scheme to that of A is B. We can clearly see that both graphs are not connected i.e. there are subgraphs which do not have any arrow connecting them. Some of these subgraphs are larger while some others contain only two stakeholders (the minimum number in such a subgraph). The following

notation is used: Role-Insource/Outsource-Onshore/Offshore, and these are represented by the acronyms presented in Section 3. A case with only two stakeholders is the subgraph with only PM-IN-ON (4) and TPM-IN-ON (7) in the actual situation. In the ideal situation there is a similar case with TPM-IN-ON (1) and SPM-IN-OFF (1). Another interesting remark is that there are stakeholders which are “hubs”, in the sense that they are closer to several others. Such a case is ARCH-IN-OFF (1) for both situations (actual and ideal).

This type of graphs allow us to identify individuals with similar views, which may be helpful when forming teams or discussing the future development of the product. Especially for teams working remotely, managerial decisions regarding the communication between individuals or groups can be benefited from such an analysis. For example, actions enhancing the communication and interaction among employees with divergent goals can diagnose and solve many cooperation problems at the early stages of a project. Such a problem could be the “isolation” of certain individuals who fail to see the “big picture” and the overall goals of a project.

6. DISCUSSION – THREATS TO VALIDITY

The illustration of the statistical framework comes with some specific threats with respect to the actual case, and hence the results from the case cannot be expected to be generalizable; they are only an illustration of the type of findings that can be obtained when using the proposed framework for this type of analysis.

6.1 Discussion of the case study results

The clustering showed that there exist two major clusters in the actual situation and the ideal situation. We have shown that the way stakeholders are clustered is not independent from the situation, i.e. there is a large body of common stakeholders that participate in a similar clustering scheme in both of the situations, actual and ideal. This result is consistent with the “garbage can model”, which found that managers prefer to address issues with incremental change, rather than rapid evolution (i.e. people are conservative when it comes to changing things) (Cohen et al., 1972).

The first cluster in both situations is the largest (for convenience we refer to the largest cluster of each scheme as Cluster 1) and is characterized by a more uniform prioritization of goals. The second cluster is much smaller than the first (its size is less than the half size of the first) and is characterized by large variations either within the same goal or among goals. This information could be especially interesting when planning brainstorming about future ideas and perspectives. Once the analysis identifies Cluster 1 as more conservative and Cluster 2 more open to new ideas on changes, probably it is wiser to put people in Cluster 2 together to discuss instead of having them as individuals within a number of people

from Cluster 1. This policy can potentially advance the opinion of a minority with good ideas that otherwise may not be heard. This is an indicative example of how the framework could help identifying individuals for different purposes.

The goals under the Features level seem to be in general most important for Cluster 2 while the goals of the System Properties level seem to be more important for Cluster 1, although some of the results are different for the ideal perspective. Also, Cluster 2 gives significantly larger importance to Time of Project Management for the actual situation.

The clustering does not seem to be significantly affected by the role of the stakeholder or the insource/outsourcing grouping. The fact that outsourcing stakeholders are not very close to each other but they are all clustered with the insource stakeholders is a positive indication for the company since it shows that this group is mixed with the internal employees and it has not the behavior of a minority which is isolated from the other internal employees. The groups defined by onshore insourcing and onshore outsourcing, work in a co-located environment. Thus it is possible for them to easily ask questions, hold meetings and have impromptu discussions. These factors are possibly the reason for the groups' alignment. On the other hand, it was found that the clustering is associated with the onshore/offshore grouping as almost all stakeholders of the offshore group are clustered together. The latter is something that high-level management must take into consideration, since it may result in misunderstandings between teams located in different sites.

Taking into consideration the correlation results among the different groups of stakeholders, insource stakeholders seem to have strong consistency in their views between the two perspectives actual and ideal for goals related to Features and System Properties, while they are not consistent for issues regarding Time and Cost. It is remarkable that regarding the role grouping, the level of alignment between SPM stakeholders is high in the majority of the goals between the two perspectives, while ARCH stakeholders show a tendency for inconsistency in almost all goals.

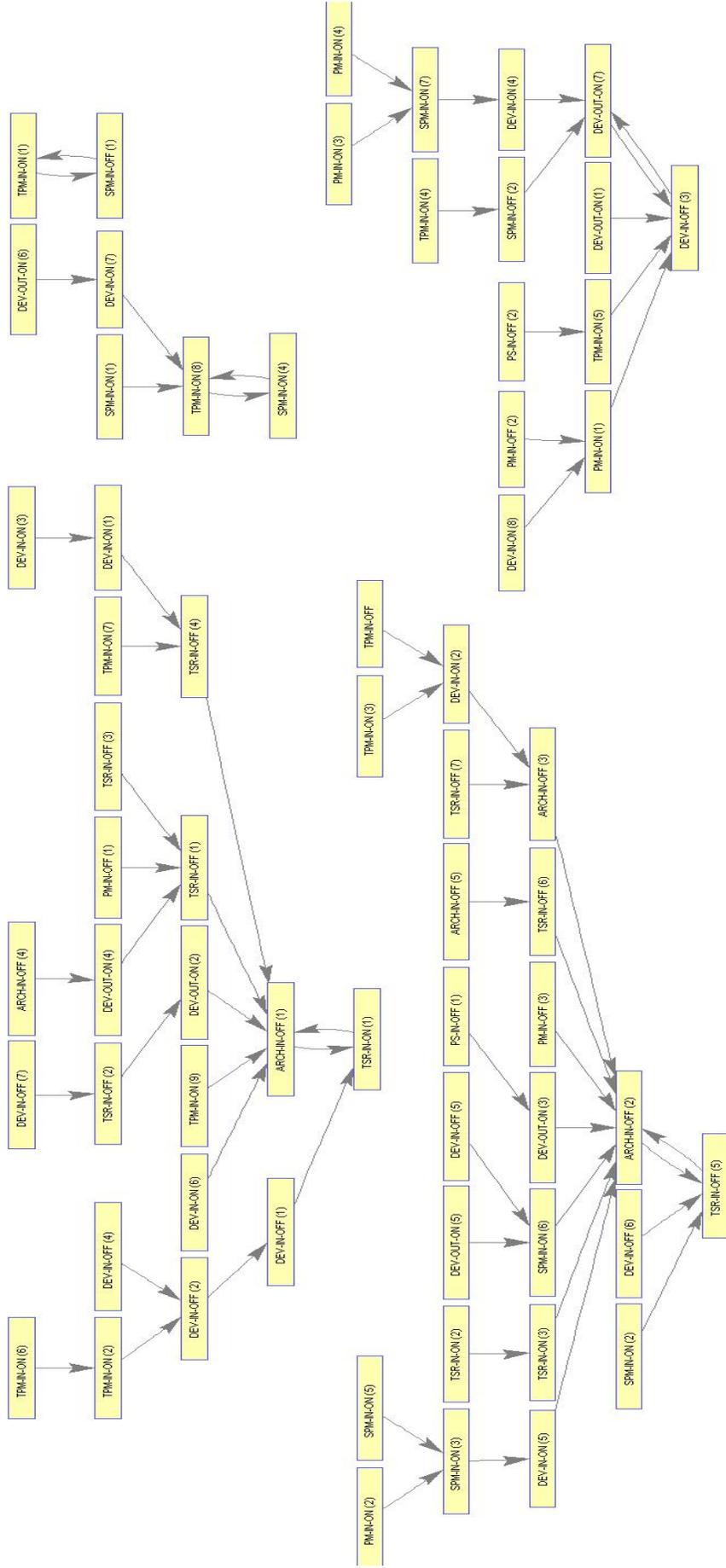


Figure 6. Closest alignment approach for ideal situation

6.2 Practical usefulness of the statistical framework

The statistical framework has been formulated based on the experiences of conducting several case studies in relation to the challenges of aligning the views and priorities of goals between different stakeholders. The statistical framework is empirically driven from the case studies conducted in industry. Thus, the actual framework as such has not been applied, although analysis of the data has been conducted in all case studies. The company perceived the analysis as valuable to them. The studies were initiated based on a perception that different roles representing different departments within the company had different priorities when it came to software goals. This was perceived as a potential problem when it came to prioritizing, for example, internal goals (e.g. evolvability) vs. external goals (e.g. new features or reliability). After having conducted the analysis, it was concluded that the people in the organization were not in as much disagreement as initially perceived. The alignment analysis was appreciated, since it gave them both a common view to discuss the goals and their priorities, and it helped them to identify both agreements and disagreements. Based on the analysis conducted, some specific actions were taken to address some goals and it also generated internal discussions that helped people getting a better understanding of the viewpoints of others.

The general conclusion from the analysis is that the methodology proposed can reveal stakeholders who agree in their prioritization and can form groups of homogeneous views. However, there are factors behind the alignment of stakeholders that remain unseen. The effect of onshore/offshore grouping on the clustering is perhaps an indication that cultural and educational characteristics are very important and that should be recorded and analyzed more. The correlation of the priorities given to the two perspectives, actual and ideal, seems to differ among the different groups of stakeholders and deserves further investigation.

The methodology presented in this paper is useful for investigating the similarities between stakeholders and the factors that may be responsible for them. Further it may be used as a tool for composing teams of professionals that share similar views on goals and have different roles or belong to different groups internal or external to a company, even from different countries. It is important to remark that each stakeholder's prioritizations, as expressed through the voting scheme employed in the study, are not disclosed to other participating stakeholders. This is an important feature of the proposed framework since it essentially eliminates any tendency for disputes or artificial disagreements originating from personal conflicts or even "politics". So the Hierarchical Cluster Analysis and the subsequent analysis of the composition of clusters and the correlation between the two perspectives can be a valuable aid to discover and bring together people that may seem completely divergent due to their positions but who in fact have similar views on critical goals.

6.3 *Validity threats*

Regarding the validity threats, we have to remark that our case study is exploratory and to a large extent illustrative of the proposed framework. It is certainly not meant to generalize the results, since the current data are not a representative sample from a population but on the contrary, the participants have been selected from a specific company and for a very specific product. The only generalization we can infer from the analysis is that if a company wants to investigate in-depth the problem of stakeholders' alignment, it needs to take into account more personal characteristics (personal skills, cultural characteristics, personality traits, etc.).

Another point that needs some clarification is the prioritization under the "ideal" perception. As we have already mentioned, the goal was to record stakeholders' prioritizations under a hypothetical and exemplary operation of the organization. Therefore this opinion was expected to be subjective and dependent on each subject's position in the organization, his or her education, expertise, culture, personality etc. It is possible that personal convenience matters could have been involved in someone's prioritization. However, our aim was not to search for the causes of the ideal prioritization, but rather to parallelize this perception with the perception of actual situation. Furthermore, it helps identifying different priorities from different individuals when it comes to where they want the organization to be heading. It makes differences and similarities visible, and hence addressable.

7. CONCLUSION

The paper focuses on a statistical framework to address issues of alignment and consistency between success-critical stakeholders involved in the development of software products. The need for agreement between various groups of stakeholders towards a common goal and under different perspectives is very important for the success of a project as emphasized by literature and shown by common practice.

The entire study presented here was originated from a very general question which we believe to be important for organizations: How can we capture and analyze in a holistic way the different views of individuals working on a project with respect to specific quality goals? This general question leads to more specific questions related to their alignment and consistency towards the goals. The statistical framework we present here is an attempt to quantify, register and statistically analyze these views, and finally to obtain a comprehensive picture of employees' opinions and attitude towards the pre-specified goals. Its development was based on the experiences gained when conducting a series of case studies in industry. Furthermore, the methods comprising the framework are quite interpretable, even by non-experts, and easily available by commercial or open/free statistical packages.

Needless to say, the results in the actual case studied when applying the methods proposed here are not expected to be the same or analogous for every organization, team or product, even within the same organization. The proposed framework should be used with awareness of the limitations that come from the data. In case the data concern certain teams and products of an organization, the findings should be considered exploratory or potentially confirmatory and should be interpreted with respect to the specific environment of the company. On the other hand, the proposed framework could be adjusted and applied to inferential studies, provided that the data will be collected as a random sample of a large population.

In this paper we chose to illustrate the proposed framework with data obtained from a single organization, knowing the limitations. However we considered the specific dataset appropriate for illustrating our research, since it comes from a field where the notion of alignment is especially important: the global software engineering field. Indeed, the employees of an organization working in a global environment for the development of a software product may have large differences in their personal characteristics depending on their personality, culture and professional or even social status. Since it is very difficult for an organization to record the views on goals of a large number of employees working on the same project and assess their agreement, the structured and systematic collection of prioritization data through the Cumulative Voting method and the multivariate statistical analysis are valuable aids for studying and understanding the reasons of alignment and possibly the causes of success or failure of a project. The case study showed that RQ01-05 can be answered by the statistical framework, which in turn shows that the framework is possible to be used on similar industrial data in relation to alignment of software goals (RQ0). Moreover, although the results should be interpreted at an organizational level, the present study showed that there is a need for collection of other attributes (for example cultural, educational, professional etc) that may reveal the latent factors that affect the alignment.

The study of connecting the prioritization alignment with alignment of other personal attributes is a direction of future research. Also another direction for future research is to go beyond the analysis results and use them to build teams, networks or to perform other managerial actions that will enhance the joint work towards common goals. Another area of research would be to apply the statistical framework after its development, and collect data about time and effort to use as well as to identify further improvements. Finally, it would be very interesting to study the automation of the whole procedure of selecting data from prioritization responses and of their statistical analysis. That could be possible by developing a software tool.

REFERENCES

- Barney, S., Wohlin, C. (2008). Software product quality questionnaire, <http://www.bth.se/tek/aps/sba>
- Barney, S., Wohlin, C. (2009a). Software product quality definitions, <http://sebseb.info/files/icsp2009-definitions.pdf>.
- Barney, S., Wohlin, C. (2009b). Software product quality: Ensuring a common goal, in Proceedings of the International Conference on Software Process (ICSP), Q. Wang R. Madachy, and D. Pfahl (Eds), pp. 256–267.
- Barney, S., Wohlin, C. (2010). Alignment of software product quality goals in two outsourcing relationships, in Proceedings of the 14th International Conference on Evaluation and Assessment in Software Engineering (EASE).
- Barney, S., Wohlin, C., Chatzipetrou, P., Angelis, L. (2011). Offshore Insourcing: A Case Study on Software Quality Alignment, in Proceedings of the 6th IEEE International Conference on Global Software Engineering, pp. 146-155.
- Barney, S., Mohankumar, V., Chatzipetrou, P., Aurum, A., Wohlin, C., & Angelis, L. (2014). Software quality across borders: Three case studies on company internal alignment. *Information and Software Technology*, 56(1), pp. 20-38.
- Berander, P, Jönsson, P. (2006a). A goal question metric based approach for efficient measurement framework definition, in: Proceedings of the 2006 ACM/IEEE International Symposium on Empirical Software Engineering (ISESE '06), pp. 316-325.
- Berander, P., Jönsson, P. (2006b). Hierarchical Cumulative Voting (HCV) – Prioritization of requirements in hierarchies, *International Journal of Software Engineering and Knowledge Engineering*, 16(6), pp. 819–849.
- Boehm, B., Ross, R. (1989). Theory-w software project management principles and examples, *IEEE Transactions on Software Engineering*, 15(7), pp. 902-916.
- Brown, S. L., Eisenhardt, K. M. (1995). Product development: past research, present findings, and future directions. *Academy of management review*, 20(2), pp. 343-378.
- Chan, Y. E. (2002). Why haven't we mastered alignment? The importance of the informal organization structure, *MIS Quarterly Executive*, 1 (2), pp. 97–112.
- Chan, Y. E., Reich, B.H. (2007). IT alignment: An annotated bibliography, *Journal of Information Technology*, 22(4), pp. 316–396.

- Chatzipetrou, P., Angelis, L., Rovegård, P., Wohlin, C. (2010). Prioritization of Issues and Requirements by Cumulative Voting: A Compositional Data Analysis Framework, in Proceedings of EUROMICRO Conference on Software Engineering and Advanced Applications (SEAA).
- Chatzipetrou, P., Angelis, L., Barney, S., Wohlin, C. (2011). Software product quality in global software development: Finding groups with aligned goals, in Proceedings of the 37th EUROMICRO Conference on. IEEE, Software Engineering and Advanced Applications (SEAA).
- Cohen, M. D., March, J. G., Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly*, 17(1), pp. 1–25.
- Conchúir, E. Ó., Ågerfalk, P. J., Olsson, H. H., Fitzgerald, B. (2009). Global software development: Where are the benefits?, *Communications of the ACM*, 52, pp. 127–131.
- Damian, D. (2007). Stakeholders in global requirements engineering: Lessons learned from practice. *Software, IEEE*, 24(2), pp. 21-27.
- Damian, D. E., Zowghi, D. (2003). RE challenges in multi-site software development organisations. *Requirements engineering*, Springer, 8(3), pp. 149-160.
- Espinosa, J. A., Slaughter, S. A., Kraut, R. E., Herbsleb, J. D. (2007). Team knowledge and coordination in geographically distributed software development, *Journal of Management Information Systems*, 24(1), pp.135-169.
- Field, A. (2009). *Discovering statistics using SPSS*, Sage Publications Limited.
- ISO 9126 (2001): Software engineering-product quality-part 1: Quality model. International Standards Organization.
- Johnson, R. A., Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis*, Prentice Hall.
- Leffingwell, D., Widrig, D. (2003). *Managing software requirements: A Use Case Approach*, 2nd ed. Addison-Wesley, Boston.
- McDonough, E. F., Spital, F. C. (2003). Managing project portfolios. *Research-Technology Management*, 46(3), pp. 40-46.
- Morgan, L. (2004). Consider the outsource. *Software Development Times*, (100), pp. 28-30.
- Pinjani, P., Palvia, P. (2013). Trust and knowledge sharing in diverse global virtual teams. *Information and Management*.
- Rovegård, P., Angelis, L., Wohlin, C. (2008). An Empirical Study on Views of Importance of Change Impact Analysis Issues, *IEEE Transactions on software engineering*, 34(4), pp. 516-530.
- Sahay, S., Nicholson, B., Krishna, S. (2003). *Global IT outsourcing: Software development across borders*, MIT Press, ISBN 0521816041.

- Sheskin, D. (2004). Handbook of Parametric and Nonparametric Statistical Procedures. Third Edition, Chapman & Hall/CRC, 2004.
- Šmite, D., Wohlin, C., Gorschek, T., Feldt, R. (2010). Empirical evidence in global software engineering: a systematic review, *Empirical Software Engineering*, 15(1), pp. 91–118.
- Šmite, D., Wohlin, C., Galviņa, Z., & Prikładnicki, R. (2014). An empirically based terminology and taxonomy for global software engineering. *Empirical Software Engineering*, 19(1), 105-153.